

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: September 10, 2003, 00:19:16 ; Search time 86 Seconds

(without alignments)  
804.707 Million cell updates/sec

Title: US-09-815-242-12600

Perfect score: 2213  
Sequence: 1 MTKPIVAIVGRPMVKSTIF.....IRAAFGECTPIHIAKRN 436

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2213	100.0	436	22	AAU37007
2	2213	100.0	436	22	AAU37139
3	2213	100.0	436	22	AAAB2089
4	2207	99.7	435	22	AAU3787
5	2094	94.6	443	23	ABP8338
6	1731	78.2	436	23	ABBA9212
7	1528	69.0	436	23	ABP7827
8	1524	68.9	319	22	AA62832
9	1523	68.8	436	23	ABP27826

10	1522	68.8	436	21	AA804108	YphC protein of St
11	1522	68.8	436	21	AA81730	Streptococcus pneu
12	1522	68.8	436	24	ABU02185	S. pneumoniae type
13	1520	68.7	436	23	ABR54068	Lactococcus lactis
14	993	44.9	448	22	AA699939	ERA binding domain
15	935.5	42.3	448	22	AA699941	ERA binding domain
16	923	41.7	503	22	AA699941	E. coli cellular p
17	907	41.0	501	22	AA699942	Salmonella typhi c
18	899	40.6	490	22	AA699942	C. glutamicum prote
19	862.5	39.0	519	22	AA699942	N. gonorrhoeae aml
20	857.5	38.7	500	24	ABP78407	Haemophilus influe
21	852	38.5	504	22	AA699945	Haemophilus influe
22	852	38.5	504	22	AA699945	Haemophilus influe
23	852	38.5	504	23	ABP5595	Bifidobacterium 10
24	851.5	38.5	461	22	AA699945	Propionibacterium
25	845.5	38.2	461	22	AA699945	Streptococcus pneu
26	841.5	38.0	446	22	AA699945	Pseudomonas aerugi
27	840.5	37.7	502	22	AA699943	ERA binding domain
28	835	37.7	502	22	AA699943	ERA binding domain
29	799.5	36.1	459	22	AA699945	Enterococcus faeca
30	776.5	35.1	459	22	AA699945	Enterococcus faeca
31	754.5	34.1	225	22	AAU33525	Amino acid sequenc
32	750.5	33.9	225	22	AAU35096	ERA binding domain
33	708.5	32.0	493	20	AAU37461	ERA binding domain
34	707	31.9	456	22	AA699937	Helicobacter pylori
35	706	31.0	458	22	AAU37461	Amino acid sequenc
36	685.5	31.0	487	20	AAU37461	ERA binding domain
37	669.5	30.3	453	22	AA699944	ERA binding domain
38	624	28.2	433	22	AA699944	ERA binding domain
39	620.5	28.0	530	22	AA699947	ERA binding domain
40	613	27.7	292	22	AA699940	ERA binding domain
41	559.5	25.3	383	17	AAU77678	ERA binding domain
42	528.5	23.9	331	23	ABU51840	ERA binding domain
43	433.5	15.4	106	22	AA699939	ERA binding domain
44	340	15.4	208	22	AA699939	S. pneumoniae derl
45	315	14.2	92	19	AA699939	

## ALIGNMENTS

RESULT 1	
ID	AAU37007 standard; Protein: 436 AA.
AC	AAU37007:
XX	14-FEB-2002 (first entry)
DT	Staphylococcus aureus cellular proliferation protein #1177.
XX	AntiSense: prokaryotic cellular proliferation protein:
DE	antibiotic: antibacterial; drug design.
XX	Staphylococcus aureus.
KW	WO200170955-A2.
XX	27-SEP-2001.
OS	21-MAR-2001; 2001WO-US099180.
PD	21-MAR-2001; 2001WO-US099180.
XX	21-MAR-2000; 2000US-191078P.
XX	23-MAY-2000; 2000US-206848P.
PR	26-MAY-2000; 2000US-207777P.
PR	23-OCT-2000; 2000US-242578P.
PR	27-NOV-2000; 2000US-253625P.
PR	22-DEC-2000; 2000US-257931P.
PR	16-FEB-2001; 2001US-269308P.
XX	(ELIT-) ELITRA PHARM INC.
PA	Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Travick JD, Carr GJ.
XX	
PI	

PI Yamamoto RT, Xu HH:

XX MPI: 2001-611495/70.

DR N-PSDB: AAS54866.

XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3: Seq ID No 12600: 511pp: English.

XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 436 AA:

Query Match 100.0%: Score 2213; DB 22: Length 436;  
Best Local Similarity 100.0%: Pred. No. 4,7e-184;  
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKRPVAVVGRPNVNGKSTIFNRIVGERSVIEDTQVTRDRIRYSSGEMLTJHDFNIIDTGC 60  
DB 1 MKRPVAVVGRPNVNGKSTIFNRIVGERSVIEDTQVTRDRIRYSSGEMLTJHDFNIIDTGC 60  
OY 61 IEIGDAPFQTOIRAQAIEAIDEADVIFVNVNREGLTQSDENVAOILYKSKKPPVLAANK 120  
DB 61 IEIGDAPFQTOIRAQAIEAIDEADVIFVNVNREGLTQSDENVAOILYKSKKPPVLAANK 120  
OY 121 VDNMEMRTDVFYSLGFGEPYRISGSHGLGDLDDAVVSHFGCEEDDPDEDTIRLSI 180  
DB 121 VDNMEMRTDVFYSLGFGEPYRISGSHGLGDLDDAVVSHFGCEEDDPDEDTIRLSI 180  
OY 181 IGRPNVNGKSLVNAIIGEDRVIVSNVAGTTRDAIDTEXSYDQODVLLIDTAGMRKKGVY 240  
DB 181 IGRPNVNGKSLVNAIIGEDRVIVSNVAGTTRDAIDTEXSYDQODVLLIDTAGMRKKGVY 240  
OY 241 ESTEKSIVLRAKAIERSNVLVYDAEGCIIEDOKRAGVANEHCKKAVVYVNMKPDVE 300  
DB 241 ESTEKSIVLRAKAIERSNVLVYDAEGCIIEDOKRAGVANEHCKKAVVYVNMKPDVE 300  
OY 301 KDSKTMMKFEDEVKRFQFLDYAQAIAFVSAKERTRLRLFPYINSESNHKKRVOSSTLN 360  
DB 301 KDSKTMMKFEDEVKRFQFLDYAQAIAFVSAKERTRLRLFPYINSESNHKKRVOSSTLN 360  
OY 361 EYVTDALSNMPTPTDGRNLNVFATOVAIIEPTTVVNVNDVLMHFYKRYLENOIRAA 420  
DB 361 EYVTDALSNMPTPTDGRNLNVFATOVAIIEPTTVVNVNDVLMHFYKRYLENOIRAA 420  
OY 421 FGEGETPIHIAAKRN 436  
DB 421 FGEGETPIHIAAKRN 436

RESULT 2  
AAU37139  
ID AAU37139 standard: Protein: 436 AA.  
XX

AC AAU37139:

XX 14-FEB-2002 (first entry)

DE *Staphylococcus aureus* cellular proliferation protein #1309.

KW Antisense: prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

OS *Staphylococcus aureus*.

PN WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001: 2001WO-US09180.

XX 21-MAR-2000: 2000US-191078P.

XX 23-MAY-2000: 2000US-206848P.

XX 26-MAY-2000: 2000US-207272P.

XX 23-OCT-2000: 2000US-242578P.

XX 27-NOV-2000: 2000US-253625P.

XX 22-DEC-2000: 2000US-257931P.

XX 16-FEB-2001: 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX MPI: 2001-611495/70.

XX N-PSDB: AAS54998.

XX New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3: Seq ID No 12732: 511pp: English.

XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 436 AA:

Query Match 100.0%: Score 2213; DB 22: Length 436;  
Best Local Similarity 100.0%: Pred. No. 4,7e-184;  
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKRPVAVVGRPNVNGKSTIFNRIVGERSVIEDTQVTRDRIRYSSGEMLTJHDFNIIDTGC 60  
DB 1 MKRPVAVVGRPNVNGKSTIFNRIVGERSVIEDTQVTRDRIRYSSGEMLTJHDFNIIDTGC 60  
OY 61 IEIGDAPFQTOIRAQAIEAIDEADVIFVNVNREGLTQSDENVAOILYKSKKPPVLAANK 120  
DB 61 IEIGDAPFQTOIRAQAIEAIDEADVIFVNVNREGLTQSDENVAOILYKSKKPPVLAANK 120  
OY 121 VDNMEMRTDVFYSLGFGEPYRISGSHGLGDLDDAVVSHFGCEEDDPDEDTIRLSI 180

DB 121 VDNMEKRTVDYDFYSLGFGEPYPSGSHGLGLDLDAVYSHFGEEDEDPYEDTIRLSI 180  
181 IGRPNVGSLSLVNAIIGEDRVYSNVAGTTRDAIDTEYSYDGDYVLTIDTACMRKKGKY 240  
DB 181 IGRPNVGSLSLVNAIIGEDRVYSNVAGTTRDAIDTEYSYDGDYVLTIDTACMRKKGKY 240  
QY 241 ESTEKSYSVLRALKAIERSNVVLYVIDAEOGIIEDDKRVAGYAHGOKAVYIVNKKDIVE 300  
DB 241 ESTEKSYSVLRALKAIERSNVVLYVIDAEOGIIEDDKRVAGYAHGOKAVYIVNKKDIVE 300  
QY 301 KDSKTKMKFEDEVRKEFOFLDYAOIAFVSAKERTRLTLPYINEASENHKKRVOSTLN 360  
DB 301 KDSKTKMKFEDEVRKEFOFLDYAOIAFVSAKERTRLTLPYINEASENHKKRVOSTLN 360  
QY 361 EYVTDATISMNPTPTDKGRRLNVFYATQVAIEPPTFVYFVNDVLMHFSYKRYLENOIRAA 420  
DB 361 EYVTDATISMNPTPTDKGRRLNVFYATQVAIEPPTFVYFVNDVLMHFSYKRYLENOIRAA 420  
QY 421 FGEGTPIHIIARKRN 436  
DB 421 FGEGTPIHIIARKRN 436  
RESULT 3  
AAB82089 standard; Protein: 436 AA.  
ID AAB82089 standard; Protein: 436 AA.  
AC AAB82089;  
XX 26-JUN-2001 (first entry)  
DT 26-JUN-2001 (first entry)  
XX Staphylococcus aureus yphc protein sequence.  
DE yphc: antimicrobial; cytostatic; antiulcer; microbial infection;  
XX gene therapy; vaccine; gastrointestinal carcinoma; gastric ulcer;  
KM gastritis.  
XX Staphylococcus aureus.  
OS Staphylococcus aureus.  
XX MO200123418-A1.  
PN 05-APR-2001.  
XX 19-SEP-2000; 2000MO-US25566.  
XX 28-SEP-1999; 99US-0406968.  
PR (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
XX Zalcain M, Blawas S, Burnham MKR, Sylvester D, Mcdevilt D;  
PI Mathle TB;  
XX MPI: 2001-308138/32.  
DR N-PSDB: AAF86461.  
XX Novel yphc polypeptides of Staphylococcus aureus useful for diagnosing  
PT and treating microbial infections, especially infection by  
PI Staphylococcus aureus and Helicobacter pylori.  
XX Claim 1; Page 3; 41pp: English.  
XX The present sequence is the yphc protein of Staphylococcus aureus. The  
CC yphc coding sequence and protein are useful for treating and diagnosing  
CC microbial infections such as infection caused by S. aureus and  
CC Helicobacter pylori. In addition, the yphc coding sequence and protein  
CC are useful for treating diseases such as H. pylori-induced cancers, e.g.  
CC gastrointestinal carcinoma, gastric ulcers, and gastritis. The present  
CC sequence was obtained from a library of clones of chromosomal DNA of  
CC S. aureus in E. coli. The sequencing data from two or more clones  
CC comprising overlapping S. aureus DNAs was used to construct the present  
CC contiguous DNA sequence.

XX SQ Sequence 436 AA;  
Query Match 100.0%; Score 2213; DB 22; Length 436;  
Best Local Similarity 100.0%; Pred. No. 4,7e-184;  
Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTRPIYAIIGRPNVGSLSLVNAIIGEDRVYSNVAGTTRDAIDTEYSYDGDYVLTIDTACMRKKGKY 60  
DB 1 MTRPIYAIIGRPNVGSLSLVNAIIGEDRVYSNVAGTTRDAIDTEYSYDGDYVLTIDTACMRKKGKY 60  
QY 61 IETGDAPFQTOIRAQAEIADIDADVIFPMVNVREGLTQSDDMAQILYKSKRPVLAIVK 120  
DB 61 IETGDAPFQTOIRAQAEIADIDADVIFPMVNVREGLTQSDDMAQILYKSKRPVLAIVK 120  
QY 121 VDNMEKRTVDYDFYSLGFGEPYPSGSHGLGLDLDAVYSHFGEEDEDPYEDTIRLSI 180  
DB 121 VDNMEKRTVDYDFYSLGFGEPYPSGSHGLGLDLDAVYSHFGEEDEDPYEDTIRLSI 180  
QY 181 IGRPNVGSLSLVNAIIGEDRVYSNVAGTTRDAIDTEYSYDGDYVLTIDTACMRKKGKY 240  
DB 181 IGRPNVGSLSLVNAIIGEDRVYSNVAGTTRDAIDTEYSYDGDYVLTIDTACMRKKGKY 240  
QY 241 ESTEKSYSVLRALKAIERSNVVLYVIDAEOGIIEDDKRVAGYAHGOKAVYIVNKKDIVE 300  
DB 241 ESTEKSYSVLRALKAIERSNVVLYVIDAEOGIIEDDKRVAGYAHGOKAVYIVNKKDIVE 300  
QY 301 KDSKTKMKFEDEVRKEFOFLDYAOIAFVSAKERTRLTLPYINEASENHKKRVOSTLN 360  
DB 301 KDSKTKMKFEDEVRKEFOFLDYAOIAFVSAKERTRLTLPYINEASENHKKRVOSTLN 360  
QY 361 EYVTDATISMNPTPTDKGRRLNVFYATQVAIEPPTFVYFVNDVLMHFSYKRYLENOIRAA 420  
DB 361 EYVTDATISMNPTPTDKGRRLNVFYATQVAIEPPTFVYFVNDVLMHFSYKRYLENOIRAA 420  
QY 421 FGEGTPIHIIARKRN 436  
DB 421 FGEGTPIHIIARKRN 436  
RESULT 4  
AAU33787 standard; Protein: 435 AA.  
ID AAU33787 standard; Protein: 435 AA.  
AC AAU33787;  
XX 14-FEB-2002 (first entry)  
DT 14-FEB-2002 (first entry)  
XX Staphylococcus aureus cellular proliferation protein #63.  
DE Antisense: prokaryotic cellular proliferation protein;  
XX antibiotic; antibacterial; drug design.  
XX Staphylococcus aureus.  
OS Staphylococcus aureus.  
XX MO200170955-A2.  
PN 27-SEP-2001.  
XX 21-MAR-2001; 2001MO-US09180.  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207127P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253525P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX (ELITR ) ELITRA PHARM INC.  
PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.  
 DR N-PSDB: AAS51646.  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 XX Example 3: Seq ID No 5283; 511bp; English.  
 XX  
 XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pcr\_sequences.  
 CC  
 XX Sequence 435 AA:  
 XX  
 XX Query Match 99.7% Score 2207; DB 22; Length 435;  
 XX Best Local Similarity 100.0%; Pred. No. 1.6e-183;  
 XX Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTKPIVAIGRPNVKSTIFNRIYGERVSIYEDPGYTRDRIRYSSGMLTHDNIIDTGC 60  
 DB 1 MTKPIVAIGRPNVKSTIFNRIYGERVSIYEDPGYTRDRIRYSSGMLTHDNIIDTGC 60  
 QY 61 IEIGDAFFQOIRAOAEIAIDADVIFNVNREGLTOSDEMVAQILYKSKKPVLAIVNK 120  
 DB 61 IEIGDAFFQOIRAOAEIAIDADVIFNVNREGLTOSDEMVAQILYKSKKPVLAIVNK 120  
 QY 121 VDNKMKRTDYDYSLGFGEPYPIGSHGGLGDLDAVVSHECEEDPYDDDTIRLSI 180  
 DB 121 VDNKMKRTDYDYSLGFGEPYPIGSHGGLGDLDAVVSHECEEDPYDDDTIRLSI 180  
 QY 181 IGRPNVKSLLVNAIIGEDRYVSNVACTRDADIDTESYDGDYVLIIDTAGRRKKGYV 240  
 DB 181 IGRPNVKSLLVNAIIGEDRYVSNVACTRDADIDTESYDGDYVLIIDTAGRRKKGYV 240  
 QY 241 ESTEKYSVLRAKAIERSNVLVYDAEOGIIEODKRVAGYAHBOGAAYIVVNNKDDTVE 300  
 DB 241 ESTEKYSVLRAKAIERSNVLVYDAEOGIIEODKRVAGYAHBOGAAYIVVNNKDDTVE 300  
 QY 301 KOSKTKMKFEDEVKREFOPLDYAOIAFVSAKERTRLTLPYINAESENKKRVOSSTLN 360  
 DB 301 KOSKTKMKFEDEVKREFOPLDYAOIAFVSAKERTRLTLPYINAESENKKRVOSSTLN 360  
 QY 361 EYVTDALISNNPTPTDKRGLNVFYATQVAIEPTFEVFNVDVLMHRSRYRYLENOIRAA 420  
 DB 361 EYVTDALISNNPTPTDKRGLNVFYATQVAIEPTFEVFNVDVLMHRSRYRYLENOIRAA 420  
 QY 421 ECEECTPIHIIARKR 435  
 DB 421 ECEECTPIHIIARKR 435

XX 24-JUL-2002 (first entry)  
 DT Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3183.  
 DE Staphylococcus epidermidis: open reading frame; ORF: bacterial infection;  
 XX Staphylococcus epidermidis: gene therapy.  
 KW antibacterial; gene therapy.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.  
 XX  
 PF 13-AUG-1998; 98US-0134001.  
 XX  
 PR 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Doucette-Stamm LA, Bush D;  
 XX  
 DR WPI: 2002-381255/41.  
 DR N-PSDB: ABN90883.  
 XX  
 PS Disclosure: SEQ ID 3183; 267bp; English.  
 XX  
 XX ABN90538 to ABN9374 represent *Staphylococcus epidermidis* open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP3124 to ABP37960. The *S. epidermidis* sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly *S. epidermidis* infections. The sequences can be used to  
 CC screen for compounds able to interfere with the *S. epidermidis* life  
 CC cycle or inhibit *S. epidermidis* infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 CC  
 XX Sequence 443 AA:  
 XX  
 XX Query Match 94.6% Score 2094; DB 22; Length 443;  
 XX Best Local Similarity 93.1%; Pred. No. 1.1e-173;  
 XX Matches 406; Conservative 18; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MTKPIVAIGRPNVKSTIFNRIYGERVSIYEDPGYTRDRIRYSSGMLTHDNIIDTGC 60  
 DB 8 MTKPIVAIGRPNVKSTIFNRIYGERVSIYEDPGYTRDRIRYSSGMLTHDNIIDTGC 67  
 QY 61 IEIGDAFFQOIRAOAEIAIDADVIFNVNREGLTOSDEMVAQILYKSKKPVLAIVNK 120  
 DB 68 IEIGDAFFQOIRAOAEIAIDADVIFNVNREGLTOSDEMVAQILYKSKKPVLAIVNK 127  
 QY 121 VDNKMKRTDYDYSLGFGEPYPIGSHGGLGDLDAVVSHECEEDPYDDDTIRLSI 180  
 DB 128 VDNLEHNDIYDYSLGFGDPYPIGSHGGLGDLDAVVSHECEEDPYDDDTIRLSI 187  
 QY 181 IGRPNVKSLLVNAIIGEDRYVSNVACTRDADIDTESYDGDYVLIIDTAGRRKKGYV 240  
 DB 188 IGRPNVKSLLVNAIIGEDRYVSNVACTRDADIDTESYDGDYVLIIDTAGRRKKGYV 247  
 QY 241 ESTEKYSVLRAKAIERSNVLVYDAEOGIIEODKRVAGYAHBOGAAYIVVNNKDDTVE 300  
 DB 248 ESTEKYSVLRAKAIERSNVLVYDAEOGIIEODKRVAGYAHBOGAAYIVVNNKDDTVE 307  
 QY 301 KOSKTKMKFEDEVKREFOPLDYAOIAFVSAKERTRLTLPYINAESENKKRVOSSTLN 360  
 DB 308 KOSKTKMKFEDEVKREFOPLDYAOIAFVSAKERTRLTLPYINAESENKKRVOSSTLN 367



OY 361 EYVDAISMNPPTDKGRRLNVEFYATOVAIPEPTFVFNVDLHMFSTRYLENQIRAA 420  
 DB 368 EYVDAISMNPPTDKGRRLNVEFYATOVAIPEPTFVFNVDLHMFSTRYLENQIRAA 427  
 OY 421 FGFEGTPHIHIAKRKN 436  
 DB 428 FGFEGTPHIHIAKRKN 443

## RESULT 6

ABBA49212 standard; Protein: 436 AA.

ABBA49212;

05-FEB-2002 (first entry)

Listeria monocytogenes protein #1916.

Anti-bacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 vitamin B12; bacterial infection; disease.

Listeria monocytogenes.

WO200177335-A2.

18-OCT-2001.

11-APR-2001; 2001WO-FR01118.

11-APR-2000; 2000FR-0004629.

(INSP) INST PASTEUR.

Buchrieser C, Frangeul L, Couve E, Rusniok C, Fajhl H, Dehoux P;  
 Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 Danile J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 Chakraborty T, Dommann E, Hain T, Berche P, Chaidit A, Durant L;  
 Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 Madueno E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;  
 Rose M, Voss H;

WPI: 2002-010914/01.

Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 and prevention of Listeria and related bacterial infections, and  
 related polypeptides

Claim 6: SEQ ID No 1917; 192pp; French.

The present invention relates to the genome sequence of Listeria  
 monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 it are useful for selecting probes and primers for detecting genes in L.  
 monocytogenes and related organisms, and for studying genetic  
 polymorphisms and other genomes. The present sequence is a protein  
 encoded by the genome sequence of the present invention. Proteins  
 expressed from the genome sequence are useful for raising specific  
 antibodies. Identification of L. monocytogenes and related organisms, and  
 for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 B12. The genome sequence and proteins encoded by it are also useful for  
 selecting compounds that regulate gene expression and cell replication  
 and modulate L. monocytogenes-related diseases. In addition, the genome  
 sequence and proteins encoded by it are useful in pharmaceutical and  
 vaccines compositions for the treatment or prevention of infections by L.  
 monocytogenes and related organisms.  
 Note: The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic format directly from WIPO  
 at ftp.wipo.int/pub/published\_pcl\_sequences.

Sequence 436 AA;

Query Match 78.28; Score 1731; DB 23; Length 436;

Best Local Similarity 73.6%; Pred. No. 4.4e-142;  
 Matches 320; Conservative 65; Mismatches 50; Indels 0; Gaps 0;

OY 1 MTKPIVAIYGRPNCKSTFNRIYGERVSEIYEDPGYTRRIYSSGMLTHDNIITDGC 60  
 DB 1 MAKPVVAIYGRPNCKSTFNRIYGERVSEIYEDPGYTRRIYSSGMLTHDNIITDGC 60  
 OY 61 IETGAPPTQIRAOAEIAIDBADYITFMVAVREGLTOSDBVAQILYSSKRPVLAVM 120  
 DB 61 IDLSDEPFLQIRAOAEIAIDBADYITFMVAVREGLTOSDBVAQILYSSKRPVLAVM 120  
 OY 121 VDNMEKRTVDYFYSLSGEGEPYPIGSHGLGLDLAAVYSHGEEDEPDYEDTIRLSI 180  
 DB 121 VDNMEKRTVDYFYSLSGEGEPYPIGSHGLGLDLAAVYSHGEEDEPDYEDTIRLSI 180  
 OY 121 VDNMEKRTVDYFYSLSGEGEPYPIGSHGLGLDLAAVYSHGEEDEPDYEDTIRLSI 180  
 DB 121 VDNMEKRTVDYFYSLSGEGEPYPIGSHGLGLDLAAVYSHGEEDEPDYEDTIRLSI 180  
 OY 181 IGRPNVKSLLNALGEDRYVSVNAGTTDADIDTEXSTGDDYVLITAGRRKKGY 240  
 DB 181 IGRPNVKSLLNALGEDRYVSVNAGTTDADIDTEXSTGDDYVLITAGRRKKGY 240  
 OY 241 ESTEKYSVLRAKAIERSNVYLVYDAEGIIEBODKRVAGYAHGOKAVIYVNMKDYVE 300  
 DB 241 ESTEKYSVLRAKAIERSNVYLVYDAEGIIEBODKRVAGYAHGOKAVIYVNMKDYVE 300  
 OY 301 KOSTKMKKREDEVRKREPOFLDYAQAFSAKEPRTLTLPYINAESEMKRKYOSTLN 360  
 DB 301 KOSTKMKKREDEVRKREPOFLDYAQAFSAKEPRTLTLPYINAESEMKRKYOSTLN 360  
 OY 361 EYVDAISMNPPTDKGRRLNVEFYATOVAIPEPTFVFNVDLHMFSTRYLENQIRAA 420  
 DB 361 EYVDAISMNPPTDKGRRLNVEFYATOVAIPEPTFVFNVDLHMFSTRYLENQIRAA 420  
 OY 421 FGFEGTPHIHIAKRKN 435  
 DB 421 FGFEGTPHIHIAKRKN 435

## RESULT 7

ABP27827 standard; Protein: 436 AA.

ABP27827;

02-JUL-2002 (first entry)

Streptococcus polypeptide SEQ ID NO 4830.

Streptococcus: GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 group A streptococcus; Streptococcus pyogenes; antibacterial;  
 anti-inflammatory; infection; vaccine; meningitis; gene therapy.

Streptococcus pyogenes.

WO200234771-A2.

02-MAY-2002.

29-OCT-2001; 2001WO-GB04789.

27-OCT-2000; 2000GB-0026333.

24-NOV-2000; 2000GB-0028727.

07-MAR-2001; 2001GB-0005640.

(CHIR-) CHIRON SPM.

(GENO-) INST GENOMA RES.

WPI: 2002-352536/38.

N-PSDB: ABN68458.

New Streptococcus protein for the treatment or prevention of infection  
 or disease caused by Streptococcus bacteria, such as meningitis, and



## RESULT 9

ABP27826

ID ABP27826 standard; Protein: 436 AA.

XX ABP27826;

DT 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 4828.

KM Streptococcus: GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KM group A streptococcus; Streptococcus pyogenes; antibacterial;

KM anti-inflammatory; Infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

PN MO200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

PA (CHIR-) CHIRON SPA.

PI (GENO-) INST GENOMIC RES.

PI Telford J, Maignani V, Margalit Ros YI, Grandi G, Fraser C;

PI Tettelin H;

PI WPI: 2002-352536/38.

PI N-PSDB: ABN68457.

PS Claim 1; Page 3645; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in the specification. The proteins have antibacterial and anti-inflammatory activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly S. agalactiae and S. pyogenes. Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins.

SQ Sequence 436 AA;

Query Match

Best Local Similarity 68.81; Score 1523; DB 23; Length 436;

Matches 289; Conservative 72; Mismatches 73; Indels 2; Gaps 2;

OY 1 MTRKIVAGRPVNGSTIFNRIVGRVSIVEDTGTDRDRYSSEGLVTHDFNIDTGG 60  
 DB 1 MVLPTVAIGRPVNGSTIFNRIVGRVSIVEDTGTDRDRYSSEGLVTHDFNIDTGG 60  
 OY 61 IETGDAPFOFOIRAOEIVIDEADVIFPMVNRBELTOSDENVQIILKSKKPVLANNK 120  
 I: |||| ||: ||:||||: ||||:||||: ||||:||||: ||||:||||: ||||:||||:

## RESULT 10

AAB04108

ID AAB04108 standard; Protein: 436 AA.

XX AAB04108;

DT 11-APR-2001 (first entry)

DE yphC protein of Streptococcus pneumoniae (GTP binding protein).

KM yphC; GTP binding protein; antibody; treatment; infection;

KM bacteremia; otitis media; conjunctivitis; pneumoniae;

KM meningitis; sinusitis; pleural empyema; endocarditis.

OS Streptococcus pneumoniae.

PN MO200068427-A1.

PF 16-NOV-2000.

PF 02-MAY-2000; 2000WO-US11894.

PF 07-MAY-1999; 99US-0307003.

PA (SMK ) SMITHKLINE BEECHAM CORP.

PA (SMK ) SMITHKLINE BEECHAM PLC.

PI Zalcacain M, Biswas S, Burnham MKR, Warren PV, Ingreham RA;

PI Chalaker AF, So CY, Holmes DJ, Warren RL, Van Horn S, Trainl CM;

PI WPI: 2000-687653/67.

PI N-PSDB: AAN54516.

PT Streptococcus pneumoniae yphC protein and DNA sequence, useful for

PT treating infections, meningitis, and bacteraemia

XX Claim 1; Page 37-38; 39pp; English.

A new nucleotide is described which encodes a 436 residue Streptococcus pneumoniae yphC (GTP binding protein). The DNA sequence can be used to transform a host cell to produce the protein and the products can be used to treat bacterial infections (especially Streptococcus pneumoniae infections, and Helicobacter pylori infections), otitis media, conjunctivitis, and pneumoniae, bacteraemia, meningitis, sinusitis, pleural empyema, and endocarditis. The encoded protein and its peptide fragments can be

CC used to generate antibodies.

XX Sequence 436 AA:

Query Match 68.8%; Score 1522; DB 21; Length 436;  
Best Local Similarity 66.2%; Pred. No. 7.1e-124;  
Matches 290; Conservative 73; Mismatches 69; Indels 6; Gaps 3;

QY 1 MTKPIVAVGRPNVSKSTIFNRIVGRVSIVEDPTGVTDRIRYSGEMLTHDPNIDTGG 60  
DB 1 MALPTIAVGRPNVSKSTIFNRIVGRVSIVEDPTGVTDRIRYSGEMLTHDPNIDTGG 60  
QY 61 IEIGDAPFOVQIRAOAEIADADVITFVNVNREGTOSDEMVAOILYKSKRPVLAUNK 120  
DB 61 IDVDAPFMEQIKHQAIEADADVITFVNVNREGTOSDEMVAOILYKSKRPVLAUNK 120  
QY 121 VDNEMKRTDVFYSGLGCEPPYISGSHGLGDLDAVSHFGE--EEDPYDEDTIRL 178  
DB 121 VDNEMKRTDVFYSGLGCEPPYISGSHGLGDLDAVSHFGE--EEDPYDEDTIRL 178  
QY 179 SIIGRPNVGKSSLVNAILGEDRVIVSNVAGTTRDAIDTEYS-YGQGVYVLIOTAGMKRG 237  
DB 178 SIIGRPNVGKSSLVNAILGEDRVIVSNVAGTTRDAIDTEYS-YGQGVYVLIOTAGMKRG 237  
QY 238 KYESTEKYSVLRAKATERSNVVLYVIDAEGITIEDOKRVAAGAHGOKRAVYIVNWKMD 297  
DB 238 KYESTEKYSVLRAKATERSNVVLYVIDAEGITIEDOKRVAAGAHGOKRAVYIVNWKMD 297  
QY 298 TVERKSKTKMKFEDEVKKEFOFLDYAOIAFVSAKERTRLATLPFYINEASENHKKRVQSS 357  
DB 298 TVERKSKTKMKFEDEVKKEFOFLDYAOIAFVSAKERTRLATLPFYINEASENHKKRVQSS 357  
QY 358 TLENDNTMTKMKNEEDIREQFOYLPYAPITIFVSALTOKRLHKLPEMIKOISESONTTRIPSA 357  
DB 358 TLENDNTMTKMKNEEDIREQFOYLPYAPITIFVSALTOKRLHKLPEMIKOISESONTTRIPSA 357  
QY 358 TLENDNTMTKMKNEEDIREQFOYLPYAPITIFVSALTOKRLHKLPEMIKOISESONTTRIPSA 357  
DB 358 TLENDNTMTKMKNEEDIREQFOYLPYAPITIFVSALTOKRLHKLPEMIKOISESONTTRIPSA 357  
QY 418 RAAFECEGTPIHIIAKRR 435  
DB 418 RAAFECEGTPIHIIAKRR 435

RESULT 11  
AAV81730  
ID AAV81730 standard; Protein: 436 AA.

AAV81730:  
02-JUN-2000 (first entry)  
Streptococcus pneumoniae protein sequence ID36.  
Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
kidney disease; diabetes; immunosuppressive disorder; otitis media;  
pneumococcal septicemia; sinusitis; meningitis; therapy.

OS Streptococcus pneumoniae.  
XX W0200006738-A2.  
PN 10-FEB-2000.  
PD 27-JUL-1999; 99WO-GB02452.  
PE 27-JUL-1998; 98GB-0016336.  
PR 19-MAR-1999; 99US-0125329.  
XX (MICR-) MICROBIAL TECHNIQS LTD.  
PA 1e Page RMF, Wells JM, Hanniffy SB, Hansbro PM;  
PI MPI; 2000-195301/17.  
XX N-PSDB; AA291826.  
DR

XX Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections  
XX Claim 2; Page 49; 76pp; English.

This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicemia, otitis media, sinusitis, and  
CC meningitis.

Sequence 436 AA:  
Query Match 68.8%; Score 1522; DB 21; Length 436;  
Best Local Similarity 66.2%; Pred. No. 7.1e-124;  
Matches 290; Conservative 73; Mismatches 69; Indels 6; Gaps 3;

QY 1 MTKPIVAVGRPNVSKSTIFNRIVGRVSIVEDPTGVTDRIRYSGEMLTHDPNIDTGG 60  
DB 1 MALPTIAVGRPNVSKSTIFNRIVGRVSIVEDPTGVTDRIRYSGEMLTHDPNIDTGG 60  
QY 61 IEIGDAPFOVQIRAOAEIADADVITFVNVNREGTOSDEMVAOILYKSKRPVLAUNK 120  
DB 61 IDVDAPFMEQIKHQAIEADADVITFVNVNREGTOSDEMVAOILYKSKRPVLAUNK 120  
QY 121 VDNEMKRTDVFYSGLGCEPPYISGSHGLGDLDAVSHFGE--EEDPYDEDTIRL 178  
DB 121 VDNEMKRTDVFYSGLGCEPPYISGSHGLGDLDAVSHFGE--EEDPYDEDTIRL 178  
QY 179 SIIGRPNVGKSSLVNAILGEDRVIVSNVAGTTRDAIDTEYS-YGQGVYVLIOTAGMKRG 237  
DB 178 SIIGRPNVGKSSLVNAILGEDRVIVSNVAGTTRDAIDTEYS-YGQGVYVLIOTAGMKRG 237  
QY 238 KYESTEKYSVLRAKATERSNVVLYVIDAEGITIEDOKRVAAGAHGOKRAVYIVNWKMD 297  
DB 238 KYESTEKYSVLRAKATERSNVVLYVIDAEGITIEDOKRVAAGAHGOKRAVYIVNWKMD 297  
QY 298 TVERKSKTKMKFEDEVKKEFOFLDYAOIAFVSAKERTRLATLPFYINEASENHKKRVQSS 357  
DB 298 TVERKSKTKMKFEDEVKKEFOFLDYAOIAFVSAKERTRLATLPFYINEASENHKKRVQSS 357  
QY 358 TLENDNTMTKMKNEEDIREQFOYLPYAPITIFVSALTOKRLHKLPEMIKOISESONTTRIPSA 357  
DB 358 TLENDNTMTKMKNEEDIREQFOYLPYAPITIFVSALTOKRLHKLPEMIKOISESONTTRIPSA 357  
QY 358 TLENDNTMTKMKNEEDIREQFOYLPYAPITIFVSALTOKRLHKLPEMIKOISESONTTRIPSA 357  
DB 358 TLENDNTMTKMKNEEDIREQFOYLPYAPITIFVSALTOKRLHKLPEMIKOISESONTTRIPSA 357  
QY 418 RAAFECEGTPIHIIAKRR 435  
DB 418 RAAFECEGTPIHIIAKRR 435

RESULT 12  
ABU02185  
ID ABU02185 standard; Protein: 436 AA.  
XX ABU02185:  
XX 11-FEB-2003 (first entry)  
XX S. pneumoniae type 4 strain protein from coding region #1762.  
XX Bacterial meningitis; pneumonia; sepsis; otitis media;  
KW

ear infection; antiinflammatory; antibacterial; immunostimulant;  
 auditory; respiratory; gene therapy; vaccine.  
 Streptococcus pneumoniae type 4 strain.  
 WO200277021-A2.  
 03-OCT-2002.  
 27-MAR-2002: 2002MO-IB02163.  
 27-MAR-2001: 2001GB-0007658.  
 (CHIR-) CHIRON SPA.  
 (GENO-) INST GENOMIC RES.  
 Masignani V, Tettelin H, Fraser C;  
 NPI: 2003-040579/03.  
 N-PSDB: ABX07474.  
 New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 useful as medicaments for treating or preventing a disease or infection  
 due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
 or ear infection  
 Claim 1: SEQ ID NO 3524: 56pp: English.  
 The invention relates to a protein comprising or having at least 508  
 amino acid residues, identified in the  
 specification (available on a computer readable format), or its fragment,  
 expressed from 2469 identified DNA coding regions from the  
 Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 AB556454. Also included are an antibody which binds one of the  
 proteins, treating a patient by administering the protein, DNA or  
 antibody (in a composition), a kit comprising first and second primers,  
 which are the nucleic acid cited above or fragments between nucleotides  
 8-100 of a sequence not defined in the specification, for amplifying a  
 target sequence contained within a Streptococcus nucleic acid sequence,  
 where the first primer is substantially complementary to the target  
 sequence and the second primer is substantially complementary to the  
 complement of the target sequence, and where the parts of the primers  
 having substantial complementarity define the termini of the target  
 sequence to be amplified, assay comprising contacting a test compound  
 with the protein, and determining whether the test compound binds to the  
 protein and a Streptococcus pneumoniae bacterium, where one or more  
 genes encoding the proteins has been rendered inactive. The proteins,  
 nucleic acid molecules, antibody and compositions are useful as  
 medicaments for treating or preventing a disease or infection due to  
 streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,  
 sepsis, otitis media or ear infection. They are also useful in developing  
 vaccines, diagnostics and antibodies. The methods are useful for  
 identifying immunodominant proteins. The present sequence is one of  
 the 2469 proteins expressed by the identified coding regions from the  
 genomic sequence.  
 Note: The sequence data for this patent did not form part  
 of the printed specification, but was obtained in electronic  
 format directly from Wipo at  
 ftp.wipo.int/pub/published\_pct\_sequences.  
 Sequence 436 AA:  
 Query Match 68.8%: Score 1522; DB 24; Length 436;  
 Best Local Similarity 66.2%: Pred. No. 7.1e-124;  
 Matches 290: Conservative 73; Mismatches 69; Indels 6; Gaps 3;  
 OY 1 MTRPIAIVGRPNVSKSTIFNIVGRVSIYEDTGVYRDRITSSGELTDFDITDGG 60  
 DB 1 MALPTAIYGRPNVSKSTIFNIVGRVSIYEDTGVYRDRITSSGELTDFDITDGG 60  
 OY 61 IELGDAPFOTQIRAOAEIADADVIYIRNVRKSLDSDKMAQILYKSKRPVLAANK 120  
 DB 61 IDVDAPFMDQIRKQAEIADADVIYIRNVRKSLDSDKMAQILYKSKRPVLAANK 120

OY 121 VDNEMHRTVDYDFSLGCEPPEYISGSHGLGLDLDAVSHFCE--EEDPYDEDTIRL 178  
 DB 121 VDNEMHRTVDYDFSLGCEPPEYISGSHGLGLDLDAVSHFCE--EEDPYDEDTIRL 178  
 OY 179 SIIGRPVNGKSSLVNMGEDRIYVSNVAGTTRDAIDTREFS-YOCODVYLLDTGAKKKG 237  
 DB 178 SLIGRPVNGKSSLVNMGEDRIYVSNVAGTTRDAIDTREFS-YOCODVYLLDTGAKKKG 237  
 OY 238 KYESTERYVLAALKAIERSNVLYVDAEGGIIEDOKRVAGVAHEGKRAVYIVKMD 297  
 DB 238 KYESTERYVLAALKAIERSNVLYVDAEGGIIEDOKRVAGVAHEGKRAVYIVKMD 297  
 OY 298 TVEKDSKTMKFEDEYKREFQFLDYQIAFVSAKERTRLTLPYINASENHKKRQSS 357  
 DB 298 TLEKDNHTMKMEEDIEDQFQYLPVAPITFVSALTKORHLKPEIKINISSEOTRIPSA 357  
 OY 358 TLNEVYTDALSMNPTPDKGRNLNVEYATQVAIEPTTYVYNDVELMHFSYKRLENOJ 417  
 DB 358 VLNDVINDALAINPTPDKGRNLNVEYATQVAIEPTTYVYNDVELMHFSYKRLENOJ 417  
 OY 418 RAARFEGSTPIHIIARRK 435  
 DB 418 RAARFEGSTPIHIIARRK 435  
 RESULT 13  
 ID AB854068 standard; Protein: 436 AA.  
 AC AB854068;  
 DT 16-MAY-2002 (first entry)  
 DE Lactococcus lactis protein yphL.  
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
 OS Lactococcus lactis IL1403.  
 XX FR2807446-A1.  
 XX 12-OCT-2001.  
 PD 11-APR-2000; 2000FR-0004630.  
 PF 11-APR-2000; 2000FR-0004630.  
 PR 11-APR-2000; 2000FR-0004630.  
 PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
 XX Boloctine A, Sorokline A, Renault P, Ehrlich SD;  
 PI WPI: 2002-043418/06.  
 DR New nucleotide sequence useful in the identification or Lactococcus  
 PT lactis and related species -  
 PS Claim 6: SEQ ID NO 770: 2504pp: French.  
 The present invention is related to a Lactococcus lactis nucleotide  
 sequence (AB853300-AB853521). The  
 nucleic acid sequence is useful in the detection and/or amplification of  
 related species. The proteins of the invention are useful for the  
 biosynthesis or biodegradation of a composition of interest. The  
 invention helps research in lactic bacteria, particularly useful in the  
 production of yogurt and cheese.  
 Note: The sequence data for this patent is based on equivalent patent  
 WO200177334 (published 18-OCT-2001) which is available in electronic  
 format directly from Wipo at ftp.wipo.int/pub/published\_pct\_sequences.  
 Sequence 436 AA:

Query Match	68.7%	Score 1520;	DB 23;	Length 436;
Best Local Similarity	66.3%;	Pred. No. 1.1e-123;		
Matches 289;	Conservative	69;	Mismatches	76;
			Indels	2;
			Gaps	2;

[illegible]

XX	RESULT 14
XX	AAG99939
ID	AAG99939 standard; Protein; 448 AA.
XX	
AC	AAG99939:
DT	27-SEP-2001 (first entry)
DE	
XX	ERA binding domain polypeptide SEQ ID NO 381.
XX	
KM	ERA binding domain; <i>Escherichia coli</i> ; GTPase; antimicrobial;
KM	antibacterial; antibiotic; pathogenesis; Infection; vaccine;
KM	peptide therapy.
XX	
OS	<i>Mycobacterium leprae</i> .
XX	
PN	WO200153458-A2.
PD	
XX	26-JUL-2001.
PF	
XX	17-JAN-2001; 2001WO-US01786.
PR	
XX	18-JAN-2000; 2000US-0176870.
PA	
XX	(SMIK ) SMITHKLINE BEECHAM CORP.
PA	(SMIK ) SMITHKLINE BEECHAM PLC.
XX	
PI	
XX	Lupas AN, Pearce KH;
DR	
XX	WPI; 2001-476108/51.
PT	
XX	New ERA binding domain polypeptides and polynucleotides encoding them
PT	useful as research reagents and materials for discovery of treatments
PT	and diagnostics for diseases, or for genetic immunisation -
XX	
OS	Claim 1: Page 52-53: 279pp: English.

XX The present invention relates to ERA binding domain polypeptides  
CC (AMG99555-AMG99989 and AM00010-AM00002). The era gene in *Escherichia*  
CC coli codes for an essential GTPase protein able to autophosphorylate at  
CC serine and/or threonine residues. The protein has potential antimicrobial  
CC and antibacterial activity and is useful in screening for antagonists,  
CC agonists and for compounds with antibiotic activity. The proteins are  
CC also useful in determining their role in pathogenesis of infection,  
CC dysfunction and disease and could be used as part of a vaccine and/or  
CC peptide therapy.  
XX  
SQ Sequence 448 AA:

[illegible]

RESULT	15
AC	AMG99941
AC	standard; Protein: 448 AA.
AC	AMG99941;
DT	27-SEP-2001 (first entry)
DE	ERA binding domain polypeptide SEQ ID NO 383.
XX	
XX	ERA binding domain; Escherichia coli; GTPase; antimicrobial;
KM	antibacterial; antibiotic; pathogenesis; infection; vaccine;
XX	peptide therapy.
XX	
OS	Treponema pallidum.
XX	
PN	MO200153458-A2.
XX	
PD	26-JUL-2001.
XX	
PF	17-JAN-2001; 2001WO-US01786.
XX	
PR	18-JAN-2000; 2000US-0176870.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

## OM protein - protein search, using sw model

Run on: September 10, 2003, 00:29:51; Search time 43 Seconds

(without alignments)  
975.106 Million cell updates/sec

Title: US-09-815-242-12600

Perfect score: 2213  
Sequence: 1 MTKPIVAIVGRPNVCKSTIF.....IRAAFGFGCTPIHIAKRN 436Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96166682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database:

1: PIR\_76:\*  
2: PIR\_76:\*  
3: PIR\_76:\*  
4: PIR\_76:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2213	100.0	436	2 C89926	hypothetical prote
2	1731	78.2	436	2 A11316	hypothetical prote
3	1724	77.9	436	2 A11688	hypothetical prote
4	1699	76.8	436	2 A69336	probable GTP bindi
5	1658	74.9	437	2 F83854	hypothetical prote
6	1522	68.8	436	2 B95199	phosphoglycerate d
7	1520	68.7	436	2 H98065	hypothetical prote
8	1218.5	55.1	438	2 C86719	GTP-binding protei
9	1218.5	55.1	438	2 B97111	GTP-binding protei
10	1037	46.9	453	2 A81867	GTP-binding protei
11	1021	46.1	452	2 S75645	probable GTP bindi
12	995	45.0	442	2 B82899	conserved hypotet
13	984.5	44.5	442	2 H90522	GTP-binding protei
14	935.5	42.5	448	2 D64236	hypothetical prote
15	923	41.7	503	2 E91050	probable GTP-bind
16	923	41.7	503	2 A85895	probable GTP-bind
17	923	41.7	503	2 F65027	probable GTPase/GT
18	910	41.1	490	2 AF0821	probable GTP-bind
19	906.5	41.0	494	2 B82284	GTP-binding protei
20	896.5	40.5	495	2 AC0350	probable GTP-bind
21	864.5	39.1	433	2 A70465	probable GTP bindi
22	860.5	38.9	485	2 C81149	essential GTPase N
23	858.5	38.8	485	2 F81871	probable GTP-bind
24	852	38.5	504	2 F64143	conserved hypotet
25	842.5	38.1	439	2 B72253	conserved hypotet
26	840.5	38.0	493	2 B83171	GTP-binding protei
27	838	37.9	588	2 D87454	probable GTP-bind
28	806.5	36.4	463	2 H70504	GTP-binding protei
29	797.5	36.0	483	2 AH3445	GTP-binding protei

30	789	35.7	476	2 A97636	probable GTP-bind
31	789	35.7	476	2 AC2859	GTP-binding protei
32	784	35.4	438	2 F75290	conserved hypotet
33	782.5	35.4	461	2 S72953	probable GTP-bind
34	782.5	35.4	462	2 F87080	probable GTP-bind
35	779	35.2	465	2 D82804	GTP-binding protei
36	778.5	35.2	475	2 A95982	probable GTP-bind
37	752	34.0	453	2 C85000	hypothetical GTP-b
38	746.5	33.7	447	2 F97828	conserved GTP-bind
39	737.5	33.3	447	2 H71673	hypothetical prote
40	734.5	33.2	460	2 H81381	probable GTP-bind
41	719.5	32.5	490	2 A81744	GTP-binding protei
42	710.5	32.1	490	2 H71480	probable GTPase/GT
43	706	31.9	458	2 B64624	GTP-binding protei
44	701	31.7	462	2 E71891	probable GTP bindi
45	685.5	31.0	487	2 B86596	GTPase/GTP-binding

## ALIGNMENTS

## RESULT 1

C89926 hypothetical protein SA1307 [Imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: C89926

R:Kuroda, M.; Onita, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O

ma, A.; Mizutani, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hirata, K.

C.: Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hirata, K.

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: AB9758; MUID:21311952; PMID:11418146

A:Accession: C89926

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-436 &lt;KUR&gt;

A:Cross-references: GB:BA000018; PID:913701274; PIDN:BA842568.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA1307

C:Superfamily: Mycobacterium leprae probable GTP-binding protein; translation elongat

Query Match	Score	DB 2	Length	436
Best local similarity	100.0%	100.0%	Pred. No. 1e-128	
Matches	436	Conservative	0	Mismatches
			0	Indels
			0	Gaps
			0	
OY	1	MTKPIVAIVGRPNVCKSTIFNRIVGERVSTVEDTPCVTRDRITSSGEMLTDFNITDTGG	60	
DB	1	MTKPIVAIVGRPNVCKSTIFNRIVGERVSTVEDTPCVTRDRITSSGEMLTDFNITDTGG	60	
OY	61	IEIGDAPFQOTIRAOAETIADEADVIIFVNVREGLTQSDENVAOILYKSKPPVLAANK	120	
DB	61	IEIGDAPFQOTIRAOAETIADEADVIIFVNVREGLTQSDENVAOILYKSKPPVLAANK	120	
OY	121	VNNMRTDYDPSLGGEPYISGSHGAGLDDAVVSHFGSEEDPDEDTIRLSI	180	
DB	121	VNNMRTDYDPSLGGEPYISGSHGAGLDDAVVSHFGSEEDPDEDTIRLSI	180	
OY	181	ICRPVNGKSSLVNAILCEADRVYSNVAGTTRDAIDTESYDGDVYLIDTAGMRKRGVY	240	
DB	181	ICRPVNGKSSLVNAILCEADRVYSNVAGTTRDAIDTESYDGDVYLIDTAGMRKRGVY	240	
OY	241	ESTEKYSYLRALKIERSNVVLYVIDAEGCIIEQDRVAGVYHEDGKAVIVVNMKDPVE	300	
DB	241	ESTEKYSYLRALKIERSNVVLYVIDAEGCIIEQDRVAGVYHEDGKAVIVVNMKDPVE	300	
OY	301	KQSKTKMKFFEDVKKERQFLDYAOIAVSKKERTRLTLPYINASEHKKRVOSSLYLN	360	
DB	301	KQSKTKMKFFEDVKKERQFLDYAOIAVSKKERTRLTLPYINASEHKKRVOSSLYLN	360	
OY	361	EYVTAIAISMNPTPTDKGRRLNVFANOVALEPPTFVFNDELHMFYSKRYLENDIRAA	420	
DB	361	EYVTAIAISMNPTPTDKGRRLNVFANOVALEPPTFVFNDELHMFYSKRYLENDIRAA	420	



Db 361 EVYTDALISMNPTDCKGRNLNFAATQVAIEPTFFVFNVDVLMHFSYKRYLENOIRAA 420  
 Oy 421 FGECTPIHIITARRKN 436  
 Db 421 FGECTPIHIITARRKN 436

## RESULT 2

hypotheetical protein lmo1937 [Imported] - listeria monocytogenes (strain EGD-e)  
 A:Accession: A11316

C:Species: Listeria monocytogenes  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

R:Classer, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Fsihl, H.; Science 294, 849-852, 2001

A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltournam, A.; Ok, C.; Schleuter, T.; Simoes, N.; Tjere, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: A11316  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-436 <GLA>  
 A:Cross-references: GB:NC\_003210; PIDN:CAD00015.1; PID:g16411390; GSPDB:GN00177  
 A:Experimental source: strain EGD-e  
 C:Genetics:  
 A:Gene: lmo1937  
 C:Superfamily: Mycobacterium leprae probable GTP-binding protein; translation elongation

Query Match 78.2% Score 1731; DB 2: Length 436;  
 Best Local Similarity 73.6% Pred. No. 4.3e-99;  
 Matches 320; Conservative 65; Mismatches 50; Indels 0; Gaps 0;

Oy 1 MKRPVAIVGRPNVSKSTIFNRIVGERVSTVEDTPTGTRDRISGSEMLTHDFNIIDTGC 60  
 Db 1 MAKPPVAIVGRPNVSKSTIFNRIVGERVSTVEDTPTGTRDRISGSEMLTHDFNIIDTGC 60  
 Oy 61 IEIGDAPFOTOIRAOAEIAIDADVIFPMVNRGLTOSDEMVAOILYKSKRPVLAANK 120  
 Db 61 IDLSDEPLEOIRAOAEIAIDADVIFPMVNRGLTOSDEMVAOILYKSKRPVLAANK 120  
 Oy 121 VDNMEKRTDVPYFSLGCEPPTSGSHGLDGLDLDVAVSHGEEDEPTEDETRLSI 180  
 Db 121 VDNMEKRTDVPYFSLGCEPPTSGSHGLDGLDLDVAVSHGEEDEPTEDETRLSI 180  
 Oy 181 IGRPNVGSLLVNAIILGDRIVSNVAGTTRDAIDTETSYDGDVYLIDTAGMRKRGVY 240  
 Db 181 IGRPNVGSLLVNAIILGDRIVSNVAGTTRDAIDTETSYDGDVYLIDTAGMRKRGVY 240  
 Oy 241 ESTEKYSVLRAKALERSNVLVYIDAEOGIEODKRVAGYAHEOGKAVIYVNNMPDVE 300  
 Db 241 ESTEKYSVLRAKALERSNVLVYIDAEOGIEODKRVAGYAHEOGKAVIYVNNMPDVE 300  
 Oy 301 KDSKTKMKFEDEVEKEEOPFLDYAOIAVSAKERTRLTLPYINSESEHKKRVOSSTLN 360  
 Db 301 KDSKTKMKFEDEVEKEEOPFLDYAOIAVSAKERTRLTLPYINSESEHKKRVOSSTLN 360  
 Oy 361 EVYTDALISMNPTDCKGRNLNFAATQVAIEPTFFVFNVDVLMHFSYKRYLENOIRAA 420  
 Db 361 EVYTDALISMNPTDCKGRNLNFAATQVAIEPTFFVFNVDVLMHFSYKRYLENOIRAA 420  
 Oy 421 FGECTPIHIITARRK 435  
 Db 421 FGECTPIHIITARRK 435

## RESULT 3

hypotheetical protein homolog lln2051 [Imported] - listeria innocua (strain Clp11262)  
 A:Accession: A11688

C:Species: Listeria innocua  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

data no good

C:Accession: A11688  
 R:Classer, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Fsihl, H.; Science 294, 849-852, 2001  
 A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltournam, A.; Ok, C.; Schleuter, T.; Simoes, N.; Tjere, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: A11688  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-436 <GLA>  
 A:Cross-references: GB:AL592022; PIDN:CAC97281.1; PID:g16414552; GSPDB:GN00178  
 A:Experimental source: strain Clp11262  
 C:Genetics:  
 A:Gene: lln2051  
 C:Superfamily: Mycobacterium leprae probable GTP-binding protein; translation elongation

Query Match 77.9% Score 1724; DB 2: Length 436;  
 Best Local Similarity 73.1% Pred. No. 1.2e-98;  
 Matches 318; Conservative 67; Mismatches 50; Indels 0; Gaps 0;

Oy 1 MKRPVAIVGRPNVSKSTIFNRIVGERVSTVEDTPTGTRDRISGSEMLTHDFNIIDTGC 60  
 Db 1 MAKPPVAIVGRPNVSKSTIFNRIVGERVSTVEDTPTGTRDRISGSEMLTHDFNIIDTGC 60  
 Oy 61 IEIGDAPFOTOIRAOAEIAIDADVIFPMVNRGLTOSDEMVAOILYKSKRPVLAANK 120  
 Db 61 IDLSDEPLEOIRAOAEIAIDADVIFPMVNRGLTOSDEMVAOILYKSKRPVLAANK 120  
 Oy 121 VDNMEKRTDVPYFSLGCEPPTSGSHGLDGLDLDVAVSHGEEDEPTEDETRLSI 180  
 Db 121 VDNMEKRTDVPYFSLGCEPPTSGSHGLDGLDLDVAVSHGEEDEPTEDETRLSI 180  
 Oy 181 IGRPNVGSLLVNAIILGDRIVSNVAGTTRDAIDTETSYDGDVYLIDTAGMRKRGVY 240  
 Db 181 IGRPNVGSLLVNAIILGDRIVSNVAGTTRDAIDTETSYDGDVYLIDTAGMRKRGVY 240  
 Oy 241 ESTEKYSVLRAKALERSNVLVYIDAEOGIEODKRVAGYAHEOGKAVIYVNNMPDVE 300  
 Db 241 ESTEKYSVLRAKALERSNVLVYIDAEOGIEODKRVAGYAHEOGKAVIYVNNMPDVE 300  
 Oy 301 KDSKTKMKFEDEVEKEEOPFLDYAOIAVSAKERTRLTLPYINSESEHKKRVOSSTLN 360  
 Db 301 KDSKTKMKFEDEVEKEEOPFLDYAOIAVSAKERTRLTLPYINSESEHKKRVOSSTLN 360  
 Oy 361 EVYTDALISMNPTDCKGRNLNFAATQVAIEPTFFVFNVDVLMHFSYKRYLENOIRAA 420  
 Db 361 EVYTDALISMNPTDCKGRNLNFAATQVAIEPTFFVFNVDVLMHFSYKRYLENOIRAA 420  
 Oy 421 FGECTPIHIITARRK 435  
 Db 421 FGECTPIHIITARRK 435

## RESULT 4

probable GTP binding protein yphc - Bacillus subtilis  
 A:Accession: A69936

C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 02-Feb-2001

R:Kunst, F.; Ogasawara, N.; Mosser, I.; Albertini, A.M.; Alloul, G.; Azevedo, V.; Ber, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Galizzi, J.; Harwood, C.R.; Henault, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurlita, K.; Lapidus, A.; Lardino, A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau, Y.; M. Ogasawa, K.; Ogata, A.; Oudega, B.; Parv, S.H.; Parro, V.; Pohl, T.M.; Portete, Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, T.; Sato, T.; Scani, A.; Authors: Schleicher, S.; Schroeter, R.; Scollone, F.; Sekiguchi, J.; Sekowska, A.; Se

**RESULT 5**  
**F83854**  
 hypothetical protein BH1618 [imported] - *Bacillus halodurans* (strain C-125)  
 C:Species: *Bacillus halodurans*  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: F83854  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
 A:Reference number: AB9650; MUID:20512582; PMID:11058132  
 A:Accession: F83854  
 A:Status: preliminary

1 MTKPIVAIVCRPNVGKSTIFNRIVGERSIVEDTPGVTRDRIYSSGEMLTDPENIIDTGG 60

[illegible]

RESULT 7  
H98065  
phosphoglycerate dehydrogenase [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
C:Accession: H98065  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E  
y, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.U.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zock, C.; Baltz, R.H.; Jaskunas, S.R.;  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; M01D:2143245; PMID:11544234  
A:Accession: H98065  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1,436 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AL00357.1; PID:g15459218; GSPDB:GN00174  
C:Genetics:  
A:Gene: *sacA*  
C:Superfamily: Mycobacterium leprae probable GTP-binding protein; translation elongation

Query Match 68.8%; Score 1522; DB 2; Length 436;  
Best Local Similarity 66.2%; Pred. NO. 3e-86;  
Matches 290; Conservative 73; Mismatches 69; Indels 6; Gaps 3;

QY 1 MTKPIAIVAGRENVKSTIFNRIVGERSIYEDPGVTRDRISSEMLTHFDNIIDGG 60  
Db 1 MALPTIAIVGRNVKSTLFRNIGERSIYEDVEGTRDRIRATGEMLNRSFSDITCG 60

QY 61 IETGDAFPOTQIRAOAEIAIDADYIIFWVNRAGELTQSDENAOILYKSKRPVYLA 120  
Db 61 IDVDNAPFMEQIKHOAEIAEMEDADYIVFVSGKGIIDAQDEYVARKLYTHKRPVILA 120

QY 121 VDNMEKRTVDYVDFYSLGCEGPYPISGSHGLGGLDLDVAVSHFE--EEDEYDEDT 178  
Db 121 VDNPEKRNIDYDFYALGCEGPLPISVVGIGTGVDVAIYENLPNEEENP---DVIKF 177

QY 179 SITGRNVKSSLVNAILGEDRVYVSNAGTTRALIDTETS-IDGDYVLITDAGKRRKG 237  
Db 178 SLTGRNVKSSLVNAILGEDRVYVSNAGTTRALIDTHTDIDGEPFIMDITDAGKRRSG 237

```

OY      238  KYVESREKTSYLRAKKAIEERSNVYLVYDAEOGIEEDDKKAVAGAEHOGKAAVIVYNNKD 297
DB      238  KVELENEKTSYRAHRAKALDRSDVYLVATYNAEEGRLETRDKRIACFAHAGKCHITIVYNNKD 297
OY      298  TVEKDSKTKMKEFEDEYKKEEFOFLDYAOATAPYSAKERTLRFTLFPYINAESENHKKRVQSS 357
DB      298  TLEKDNHNTKMNNEEDIRFOFOLPAPPIIFASALTQKRLKLPKEMKQISSEOMNRTISA 357
OY      358  TLENEVYDAISNMPETDQGRRLNLFVATQVAIEPPFVFEVNDVELMHPFSYKRYLENOI 417
DB      358  VLNDVIMDAIATINPTDQGRKRLKFVATQVATKPPFVIFVNEBELMHFSYLRLENOI 417
OY      418  RAAPFEGCPRIHTIARRK 435
DB      418  RKAAFEGETPIHLIARRK 435

RESULT 8
C66719
CSP-binding protein [Imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C.Species: Lactococcus lactis subsp. lactis
C.Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C.Accession: C66719
R.Bolotin, A.; Muncher, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissensbach, J.; Eh
Genome Res. 11, 731-753, 2001
A.Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A.Reference number: AB6625; M0ID:2135186; PMID:11357471
A.Accession: C66719
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-436 <SP0>
A.Cross-references: GB:AE005176; PID:g12723672; PIDN:MAK04853.1; GSPDB:GN00146
A.Experimental source: strain IL1403
C.Genetics:
A.Gene: yphL
C.Superfamily: Mycobacterium leprae probable GTP-binding protein; translation elongat
Query Match 68.7%; Score 1520; DB 2; Length 436;
Best Local Similarity 66.3%; Pred. No. 4e-86;
Matches 289; Conservative 69; Mismatches 76; Indels 2; Gaps 2;

```

[illegible]



F:295-298/Region: GTP-binding NKXD motif  
F:330-332/Region: GTP-binding SAK/L motif #status atypical

Query Match 46.1% Score 1021, DB 2; Length 452;  
Best Local Similarity 45.9%; Pred. No. 2e-55;  
Matches 200; Conservative 94; Mismatches 138; Indels 4; Gaps 4;

QY 1 MTRPIVAIGRPVNGKSTIFNRIVGERVSIYEDTGTDRIRYSSGEMLTHTDNIIDTGG 60  
DB 1 MSLPVAIGRPVNGKSTIFNRIVGERVSIYEDTGTDRIRYSSGEMLTHTDNIIDTGG 60  
QY 61 IETG-DAPPOQIRAOAIEADAVIIFMNVNREGTOSDENVAQILYKSKRPVLA 119  
DB 61 LVFNDSFELPEIRDOALALAEKAAIFVVDGQGQGPASDEEIAQMLKQOSPVLA 120  
QY 120 KVDNMEM-RTDYDFYSLGCEPPYPIGSGHGLGDLDAVSHFEDEEDPYDEDTIRL 178  
DB 121 KCESPDOCAIQAAEFMHLGLCEPYPMASIHSGTGDLDALETLPAPOEEP-EDEEIKY 179  
QY 179 STIGRPVNGKSLVNAIGEDRVIVSNVAGTTTADITEXSYDGODVLIIDTGMKRGK 238  
DB 180 AIVGRPNVNGKSSLLNALGEGORATVPSITGTRALIDVVERKQKRLIDTIGIRKK 239  
QY 239 VESTERKYSVLAKAIEKSNVVLVIDAEGITEDDKRVAGYAHGOKAVIVVKNKMDT 298  
DB 240 VDYGAEPFGINRAFAIRADVVLVDVLDGVTEDDLKLAGRIEDGKRAVVLVINKMDA 299  
QY 299 VERKSKTKMKFEDEKREFOLDYAOIAFVSAKERTRLTLPYINASENKKRVQSSST 358  
DB 300 VERKSDYITIEHREQMARLYFMDNAEMIFVSAQICGLAVQILDCVDAQENHRRYITAV 359  
QY 359 LBNVYTDATISMNPTPTDK-GRRLVFAVATOVAIPEPFVFNVDVLMHSTKRYLENOI 417  
DB 360 IREVLBNVSHSPPTTHQKQKIVYGTOSTOPPAIALVNDPKNRNDVRYTEKOP 419  
QY 418 RAAGFECTPIHIIAR 433  
DB 420 RKQLGFGSPILRFWR 435

RESULT 12  
882899  
conserved hypothetical ATP/GTP-binding protein U0383 (imported) - Ureaplasma urealyticum  
C:Species: Ureaplasma urealyticum  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C:Accession: B82899  
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Helner, C.R.; Chen, E.Y.; Cassell, G.H.  
submitted to GenBank, February 2000  
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mit  
A:Reference number: A82870  
A:Accession: B82899  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-442 <GLA>  
A:Cross-references: GB:A8002135; GB:AF222894; NID:96899358; PIDN:AAF30793.1; GSPDB:GN001  
A:Experimental source: setovar 3; biovar 1  
C:Genetics:  
A:Gene: U0383  
A:Genetic code: SGC3  
C:Superfamily: Mycobacterium leprae probable GTP-binding protein: translation elongator  
F:175-296/Domain: translation elongation factor tu homology <EN>

Query Match 45.0% Score 995, DB 2; Length 442;  
Best Local Similarity 45.4%; Pred. No. 7.6e-54;

- Matches 137; Conservative 93; Mismatches 140; Indels 4; Gaps 4;

QY 6 VAIVGRPNVNGKSTIFNRIVGERVSIYEDTGTDRIRYSSGEMLTHTDNIIDTGGIETD 65  
DB 4 IAIIVGRPNVNGKSSLPENRILMRKRSIVDDQGVTRDRIRYDIGNMLTSSFLIDTGGIISK 63  
QY 66 APTQOTIRAOAIEADAVIIFMNVNREGTOSDENVAQIL-YKSK-RPVVLA 123  
DB 64 DTYQDNINDEOVLFAINENATITIFLVSAKDGINDDKIAIMLAKKAKKIIILVINKIES 123

QY 124 MEKRTDYDFYSLGCEPPYPIGSGHGLGDLDAVSHFEDEEDPYDEDTIRLSICR 183  
DB 124 EKYLLNGELYSFGFKFFKISAEGHIGCDLDELVDN-PIONNKERRERFICR 182  
QY 184 PNVKSSLVNAIGEDRVIVSNVAGTTTADITEXSYDGODVLIIDTGMKRGKRYEST 243  
DB 183 PNVKSSLVNIGEDRVIVNAGSTRSDIONDENYNNKTYIITDPAIRRGKIVEEV 242  
QY 244 EKYSVLAKAIEKSNVVLVIDAEGITEDDKRVAGYAHGOKAVIVVKNKDT-VERD 302  
DB 243 EKTAVLTKAIEKRSLLILVLDGSEPREDDQEVGGLAVIANIPTIILVKNKDNINK 302  
QY 303 SKTKMKFEDEKREFOLDYAOIAFVSAKERTRLTLPYINASENKKRVQSSSTLNEY 362  
DB 303 SHTMEMVKOIRSOFFYISMARIVVSAIDNKNRIHTIFEALEFVREOAMKRIATSLIN 362  
QY 363 VTOAISNPTPTDKGRRLVFAVATOVAIPEPFVFNVDVLMHSTKRYLENOIARQ 422  
DB 363 VIANANQEPPEPKGRISISIVYQSOIPTFVLKCNKPKFLHFVARIENIKRG 422  
QY 423 FEGTPIHIIARKRN 436  
DB 423 FDSVPTITLVQDN 436

RESULT 13  
H90522  
GTP-binding protein (imported) - Mycoplasma pulmonis (strain UAB CTIP)  
C:Species: Mycoplasma pulmonis  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
C:Accession: H90522  
R:Chambaud, I.; Hellwig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer, N.  
Nucleic Acids Res. 29, 2145-2153, 2001  
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p  
A:Reference number: A99512; NUID:21267165; PMID:11353084  
A:Accession: H90522  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-435 <KUR>  
A:Cross-references: GB:A1445566; PID:G14069501; PIDN:CAC13261.1; GSPDB:GN00153  
A:Experimental source: strain UAB CTIP  
C:Genetics:  
A:Gene: MYPU\_0880  
A:Genetic code: SGC3  
C:Superfamily: Mycobacterium leprae probable GTP-binding protein: translation elongat

Query Match 44.5% Score 984.5; DB 2; Length 435;  
Best Local Similarity 43.5%; Pred. No. 3.3e-53;  
Matches 187; Conservative 108; Mismatches 130; Indels 5; Gaps 3;

QY 5 IVAIVGRPNVNGKSTIFNRIVGERVSIYEDTGTDRIRYSSGEMLTHTDNIIDTGGIETG 64  
DB 9 LVAIVGRPNVNGKSTIFNRIVGERVSIYEDTGTDRIRYENIMSGKNFRITIDTGGIVS 68  
QY 65 DAPPOQIRAOAIEADAVIIFMNVNREGTOSDENVAQILYKSKRPVLA 124  
DB 65 DAPPOQIRAOAIEADAVIIFMNVNREGTOSDENVAQILYKSKRPVLA 124  
QY 69 DQFVQIRAOAIEESEILFVIDSEETSDDLIASILANSKKVLA 127  
DB 125 EMRTDYDFYSLGCEPPYPIGSGHGLGDLDAVSHFEDEEDPYDEDTIRLSICR 184  
DB 128 -AKNEDSYSLGFEDEYKISSVHGEGIGELVDKINLMPENDO--DEDLKRIALIGRP 184  
QY 185 NVKSSLVNAIGEDRVIVSNVAGTTTADITEXSYDGODVLIIDTGMKRGKRYEST 244  
DB 185 NVKSSLVNAIGEDRVIVSNVAGTTTADITEXSYDGODVLIIDTGMKRGKRYEST 244  
QY 245 KYSVLAKAIEKSNVVLVIDAEGITEDDKRVAGYAHGOKAVIVVKNKDTVERKSK 304  
DB 245 HVALMRAKSDSDSLIITIDATELSHPNARIIGTASDKKPTIIVINKDLIKKPN 304  
QY 305 TMKMKFEDEKREFOLDYAOIAFVSAKERTRLTLPYINASENKKRVQSSSTLNEY 364  
DB 305 TMKMKFEDEKREFOLDYAOIAFVSAKERTRLTLPYINASENKKRVQSSSTLNEY 364

[illegible]

DR InterPro; IPR006073; GTP1\_OBG.



DR Interpro: IPR002917; MMR\_HSR1.  
 DR Interpro: IPR005225; Small\_GTP.  
 DR Pfam: PF01926; MMR\_HSR1.1.  
 DR PRINTS: PRO0326; GTP\_OBG.  
 DR TIGRPFAMs: TIGR00650; MG442; 2.  
 DR TIGRPFAMs: TIGR00231; small\_GTP; 2.  
 DR GTP-binding: Repeat: Complete proteome.  
 KM NP-BIND: 10 17 GTP 1 (POTENTIAL).  
 FT NP-BIND: 57 61 GTP 1 (POTENTIAL).  
 FT NP-BIND: 119 122 GTP 1 (POTENTIAL).  
 FT NP-BIND: 182 189 GTP 2 (POTENTIAL).  
 FT NP-BIND: 229 233 GTP 2 (POTENTIAL).  
 FT NP-BIND: 294 297 GTP 2 (POTENTIAL).  
 SO SEQUENCE 436 AA: 48979 MW: 9544281862F229D3 CRC64;

Query Match 100.0%; Score 2213; DB 1; Length 436;  
 Best Local Similarity 100.0%; Pred. No. 2, 7e-128;  
 Matches 436; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKRPVIAVGRPNVSKSTIFNRIGERSYVEDTPGVTRDRITVSSGEMLTHTDNIIDTGC 60  
 DB 1 MKRPVIAVGRPNVSKSTIFNRIGERSYVEDTPGVTRDRITVSSGEMLTHTDNIIDTGC 60  
 OY 61 IEIGDAPFOTOIRAOAEIAIDADVIIIPVNVREGLTQSDENVAQILYKSKRPVLAANK 120  
 DB 61 IEIGDAPFOTOIRAOAEIAIDADVIIIPVNVREGLTQSDENVAQILYKSKRPVLAANK 120  
 OY 121 VONMEMRTDYDFYSLGFGEPYPISSHGGLGDLDAVYSHFGEEDEPDYEDDTIRLSI 180  
 DB 121 VONMEMRTDYDFYSLGFGEPYPISSHGGLGDLDAVYSHFGEEDEPDYEDDTIRLSI 180  
 OY 181 IGRPNVGSLLVNALGDRIVSNVAGTTRDAIDTETSYDGDYVILDTAGRRKRGKY 240  
 DB 181 IGRPNVGSLLVNALGDRIVSNVAGTTRDAIDTETSYDGDYVILDTAGRRKRGKY 240  
 OY 241 ESTEKYSVLRAKAIERSNVLVYIDAEGIIEDDKRVAGYAHDOGRAVYVNNKDYVE 300  
 DB 241 ESTEKYSVLRAKAIERSNVLVYIDAEGIIEDDKRVAGYAHDOGRAVYVNNKDYVE 300  
 OY 301 KOSKTKKFEDEVREKFEFLDYAOIAFVSAKERTLTLPYINAESENHKKRVOSTLN 360  
 DB 301 KOSKTKKFEDEVREKFEFLDYAOIAFVSAKERTLTLPYINAESENHKKRVOSTLN 360  
 OY 361 EYVTDASNPPTDKGRLLVFAVATQVAIEPPFVFNVDVDELMSFKRYLENOIRAA 420  
 DB 361 EYVTDASNPPTDKGRLLVFAVATQVAIEPPFVFNVDVDELMSFKRYLENOIRAA 420  
 OY 421 FGEGPTPIHTIARRN 436  
 DB 421 FGEGPTPIHTIARRN 436

RESULT 2  
 ENCA\_LISTMO STANDARD: PRT: 436 AA.  
 ID ENCA\_LISTMO STANDARD: PRT: 436 AA.  
 AC 08Y5W8;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE Probable GTP-binding protein enca.  
 GN ENCA OR LMO1937.  
 OS Listeria monocytogenes.  
 OC Bacteria: Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1639;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EGD-e / Serovar 1/2a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Glaeser P., Berche P., Blocher H., Brandt P., Chakraborty T.,  
 RA Charlit A., Chetoui F., Couve E., de Darvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Duranc L., Dussutget O.,  
 RA Ertlan K.-D., Fainl H., Garcia-del Portillo F., Garrido P.,

RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Krappek G.,  
 RA Madueno E., Maitouram A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Novales E., Novella S., de Padlos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,  
 RA Science 294:848-852(2001).  
 FL "Comparative genomics of Listeria species."  
 CC -I- SIMILARITY: BELONGS TO THE ERA/THRE FAMILY OF GTP-BINDING  
 CC PROTEINS. ENCA SOBFAILTY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).

DR EMBL: AL591981; CAD00015.1; -.  
 DR PIR: A1316; A1316.  
 DR ListList: LMO01937; -.  
 DR HAMAP: MF\_00195; - 1.  
 DR Interpro: IPR005289; GTP-binding\_dom.  
 DR Interpro: IPR006073; GTP\_OBG.  
 DR Interpro: IPR002917; MMR\_HSR1.  
 DR Interpro: IPR005225; Small\_GTP.  
 DR Pfam: PF01926; MMR\_HSR1.1.  
 DR PRINTS: PRO0326; GTP\_OBG.  
 DR TIGRPFAMs: TIGR00650; MG442; 2.  
 DR TIGRPFAMs: TIGR00231; small\_GTP; 2.  
 DR GTP-binding: Repeat: Complete proteome.  
 KM NP-BIND: 10 17 GTP 1 (POTENTIAL).  
 FT NP-BIND: 57 61 GTP 1 (POTENTIAL).  
 FT NP-BIND: 119 122 GTP 1 (POTENTIAL).  
 FT NP-BIND: 182 189 GTP 2 (POTENTIAL).  
 FT NP-BIND: 229 233 GTP 2 (POTENTIAL).  
 FT NP-BIND: 294 297 GTP 2 (POTENTIAL).  
 SO SEQUENCE 436 AA: 49144 MW: 58435C8701A66A3 CRC64;

Query Match 78.2%; Score 1731; DB 1; Length 436;  
 Best Local Similarity 73.6%; Pred. No. 7, 6e-99;  
 Matches 320; Conservative 65; Mismatches 50; Indels 0; Gaps 0;

OY 1 MKRPVIAVGRPNVSKSTIFNRIGERSYVEDTPGVTRDRITVSSGEMLTHTDNIIDTGC 60  
 DB 1 MKRPVIAVGRPNVSKSTIFNRIGERSYVEDTPGVTRDRITVSSGEMLTHTDNIIDTGC 60  
 OY 61 IEIGDAPFOTOIRAOAEIAIDADVIIIPVNVREGLTQSDENVAQILYKSKRPVLAANK 120  
 DB 61 IEIGDAPFOTOIRAOAEIAIDADVIIIPVNVREGLTQSDENVAQILYKSKRPVLAANK 120  
 OY 121 VONMEMRTDYDFYSLGFGEPYPISSHGGLGDLDAVYSHFGEEDEPDYEDDTIRLSI 180  
 DB 121 VONMEMRTDYDFYSLGFGEPYPISSHGGLGDLDAVYSHFGEEDEPDYEDDTIRLSI 180  
 OY 181 IGRPNVGSLLVNALGDRIVSNVAGTTRDAIDTETSYDGDYVILDTAGRRKRGKY 240  
 DB 181 IGRPNVGSLLVNALGDRIVSNVAGTTRDAIDTETSYDGDYVILDTAGRRKRGKY 240  
 OY 241 ESTEKYSVLRAKAIERSNVLVYIDAEGIIEDDKRVAGYAHDOGRAVYVNNKDYVE 300  
 DB 241 ESTEKYSVLRAKAIERSNVLVYIDAEGIIEDDKRVAGYAHDOGRAVYVNNKDYVE 300  
 OY 301 KOSKTKKFEDEVREKFEFLDYAOIAFVSAKERTLTLPYINAESENHKKRVOSTLN 360  
 DB 301 KOSKTKKFEDEVREKFEFLDYAOIAFVSAKERTLTLPYINAESENHKKRVOSTLN 360  
 OY 361 EYVTDASNPPTDKGRLLVFAVATQVAIEPPFVFNVDVDELMSFKRYLENOIRAA 420  
 DB 361 EYVTDASNPPTDKGRLLVFAVATQVAIEPPFVFNVDVDELMSFKRYLENOIRAA 420  
 OY 421 FGEGPTPIHTIARRN 436  
 DB 421 FGEGPTPIHTIARRN 436





```

RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scifone F.,
RA Sekiyuchi J., Sekowska A., Sefor S.J., Serror P., Shin B.S., Soldo B.,
RA Socolin A., Taccori E., Takagi T., Takahashi H., Takemaru K.,
RA Tokachi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Toseko V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Vardi A., Wambutt R., Medler E., Medler H., Melitzenger T.,
RA Winters P., Wipac A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zupchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis".
RL Nature 390:249-256(1997).
RN 131
RP SEQUENCE OF 185-436 FROM N.A.
RC STRAIN-168;
RX MEDLINE-96011379; PubMed-7592341;
RA Moridoni H.R., de Mendoza D., Cronan J.E. Jr.;
RT "Synthesis of an-glycerol 3-phosphate, a key precursor of membrane
RT lipids, in Bacillus subtilis".
RL J. Bacteriol. 177:5899-5905(1995).
CC -1- SIMILARITY: BELONGS TO THE ERA/TIME FAMILY OF GTP-BINDING
CC PROTEINS. ENCA SUBFAMILY.
CC -1- CAUTION: Ref.'s sequence differs from that shown due to a
CC frameshift.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L47648; AAC3966.1;
DR EMBL: Z59115; CAB14200.1;
DR EMBL: U32164; -; NOT_ANNOTATED_CDS.
DR PIR: A69936; A69936.
DR Subtilist; BG11443; engA.
DR HAMAP: MF_00195; -; 1.
DR InterPro: IPR005289; GTP-binding_dom.
DR InterPro: IPR006073; GTP1_OBG.
DR InterPro: IPR002917; MMR_HSR1.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF01926; MMR_HSR1; 1.
DR PRINTS: PR00326; GTP1_OBG.
DR TIGRfams: TIGR00650; MG442; 2.
DR TIGRfams: TIGR00231; small_GTP; 2.
KM GTP-binding: Repeat: Complete proteome.
FT NP_BIND 10 17 GTP 1 (POTENTIAL).
FT NP_BIND 57 61 GTP 1 (POTENTIAL).
FT NP_BIND 119 122 GTP 1 (POTENTIAL).
FT NP_BIND 182 189 GTP 2 (POTENTIAL).
FT NP_BIND 229 233 GTP 2 (POTENTIAL).
FT NP_BIND 294 297 GTP 2 (POTENTIAL).
FT CONFLICT 186 187 VG -> CR (in Ref. 2).
SO SEQUENCE 436 AA; 48769 MW; ASCCT028FFBSA442 CRC64;

Query Match 76.8%; score 1699; DB 1; Length 436;
Best Local Similarity 73.6%; Pred. No. 6.8e-97;
Matches 330; Conservative 57; Mismatches 58; Indels 0; gaps 0;

```

```

OY 181 IGRPNKSSLVNALLGEDRYVSNVACTTRDAIDTEXSYDGDYVLIOTACRKKKGY 240
DB 181 IGRPNKSSLVNALLGEDRYVSNVACTTRDAIDTEXSYDGDYVLIOTACRKKKGY 240
OY 241 ESTEKSVLRALKATERSNVVLVIDAEGITTEBOKRVAGYAHBOGKAVIYVNMWYVE 300
DB 241 ESTEKSVLRALKATERSNVVLVIDAEGITTEBOKRVAGYAHBOGKAVIYVNMWYVE 300
OY 301 KDKTKKKRFEDEVRKEFOFLDYAOIAFVSAREKTRLTLPPIYMEASENKKRVOSSSTLN 360
DB 301 KDKTKKKRFEDEVRKEFOFLDYAOIAFVSAREKTRLTLPPIYMEASENKKRVOSSSTLN 360
OY 361 EYVTDASINAPPTDGRRLNVEFYVAIEPFFVYFVNDVLMHSFRYRIENDIRAA 420
DB 361 EYVTDASINAPPTDGRRLNVEFYVAIEPFFVYFVNDVLMHSFRYRIENDIRAA 420
OY 421 FGEGTPIHIIARKR 435
DB 421 FGEGTPIHIIARKR 435
OY 421 FGEGTPIHIIARKR 435
DB 421 FGEGTPIHIIARKR 435

RESULT 5
ENCA_BACHD STANDARD: PRT: 437 AA.
AC Q9KCD4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable GTP-binding protein engA.
GN ENCA OR BH1638.
OS Bacillus halodurans.
CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86655;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RX MEDLINE-20512582; PubMed-11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis".
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- SIMILARITY: BELONGS TO THE ERA/TIME FAMILY OF GTP-BINDING
CC PROTEINS. ENCA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AP001512; BAB05357.1;
DR PIR: F83854; F83854.
DR HAMAP: MF_00195; -; 1.
DR InterPro: IPR005289; GTP-binding_dom.
DR InterPro: IPR006073; GTP1_OBG.
DR InterPro: IPR002917; MMR_HSR1.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF01926; MMR_HSR1; 1.
DR PRINTS: PR00326; GTP1_OBG.
DR TIGRfams: TIGR00650; MG442; 2.
DR TIGRfams: TIGR00231; small_GTP; 2.
KM GTP-binding: Repeat: Complete proteome.
FT NP_BIND 10 17 GTP 1 (POTENTIAL).
FT NP_BIND 57 61 GTP 1 (POTENTIAL).
FT NP_BIND 119 122 GTP 1 (POTENTIAL).
FT NP_BIND 182 189 GTP 2 (POTENTIAL).
FT NP_BIND 229 233 GTP 2 (POTENTIAL).
FT NP_BIND 294 297 GTP 2 (POTENTIAL).
SO SEQUENCE 437 AA; 49024 MW; CIA4BD65A3ADAB CRC64;

```

Query Match 74.9%; Score 1658; DB 1; Length 437;  
 Best Local Similarity 70.9%; Pred. No. 2,26-94;  
 Matches 309; Conservative 67; Mismatches 60; Indels 0; Gaps 0;

1 MTKPIVAVGRPNVNGKSTIFNRIYGERVSYEDPTGRTORIRYSSGEMLTGHDNITDGG 60  
 1 MKRPVAVGRPNVNGKSTIFNRIYGERVSYEDPTGRTORIRYSSGEMLTGHDNITDGG 60  
 61 IIGDAPFOTQIRAOAETIADEADYIIFPVNREGLTQSDENVAQILYKSKRPVLAANK 120  
 61 IIGDAPFOTQIRAOAETIADEADYIIFPVNREGLTQSDENVAQILYKSKRPVLAANK 120  
 61 IIGDAPFOTQIRAOAETIADEADYIIFPVNREGLTQSDENVAQILYKSKRPVLAANK 120  
 121 VDNMERTDVIDYDFYSLGFGEPYPIGSHGLGLDGLAVVSHFGEEDPDYEDTIRLSI 180  
 121 IDHPNOBELYEFYSLGFGEPYPIGSHGLGLDGLAVVSHFGEEDPDYEDTIRLSI 180  
 181 IGRPNVNGKSLVNAALGEDRVYVSNVAGTTPDAIDTGYSTODGDYVLTIDTACKRKKGY 240  
 181 IGRPNVNGKSLVNAALGEDRVYVSNVAGTTPDAIDTGYSTODGDYVLTIDTACKRKKGY 240  
 181 IGRPNVNGKSLVNAALGEDRVYVSNVAGTTPDAIDTGYSTODGDYVLTIDTACKRKKGY 240  
 241 ESTERYSVLRALKAIERSNVVLYDAEOGIIEDDKRVAGYAEHOGKAVVYVNMKDIVE 300  
 241 ESTERYSVLRALKAIERSNVVLYDAEOGIIEDDKRVAGYAEHOGKAVVYVNMKDIVE 300  
 241 ESTERYSVLRALKAIERSNVVLYDAEOGIIEDDKRVAGYAEHOGKAVVYVNMKDIVE 300  
 301 KDKTKMKFDEDEVRKPEFQIDYADAPASAKERTRLTLPYINAESENKRRVOSTLN 360  
 301 KDKTKMKFDEDEVRKPEFQIDYADAPASAKERTRLTLPYINAESENKRRVOSTLN 360  
 301 KDKTKMKFDEDEVRKPEFQIDYADAPASAKERTRLTLPYINAESENKRRVOSTLN 360  
 361 EYVTDATSNMPTPTDKGRRLNVYATOVAIEPTFVVVNDVLMHFSKRYLENOIRAA 420  
 361 EYVTDATSNMPTPTDKGRRLNVYATOVAIEPTFVVVNDVLMHFSKRYLENOIRAA 420  
 361 EYVTDATSNMPTPTDKGRRLNVYATOVAIEPTFVVVNDVLMHFSKRYLENOIRAA 420  
 421 FGEETPIHIIIAKRN 436  
 421 FGEETPIHIIIAKRN 436  
 421 FGEETPIHIIIAKRN 436

RESULT 6  
 ENCA\_STRPY STANDARD: PRT; 436 AA.

AC 09AID0;  
 DT 28-FEB-2003 (rel. 41, Created)  
 DT 28-FEB-2003 (rel. 41, Last sequence update)  
 DT 28-FEB-2003 (rel. 41, Last annotation update)  
 DE Probable GTP-binding protein enga.  
 GN ENGA OR PGDA OR SPY0341 OR SPYM18\_0334.  
 OS Streptococcus pyogenes, and  
 OS Streptococcus pyogenes (serotype M18).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus;  
 ON NCBI\_TaxID=1314, 186103;  
 OX 111  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SF370 / ATCC 700294 / Serotype M1;  
 RC MEDLINE-21192684; PubMed-11296296;  
 RA Ferritelli J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
 RA Pliamex C., Szate S., Svorov A.N., Kenton S., Lai H.S., Lin S.P.,  
 RA Qian X., Clifton S.W., Roe B.A., McLaughlin R.,  
 RA Tian X., Clifton S.W., Roe B.A., McLaughlin R.,  
 RA Complete genome sequence of an M1 strain of Streptococcus pyogenes.\*;  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
 RN 121  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MGAS8233 / Serotype M18;  
 RC MEDLINE-21927593; PubMed-11917108;  
 RA Smoot J.C., Barblan K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
 RA Sylve G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,  
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,  
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.,  
 RT Genome sequence and comparative microarray analysis of serotype M18  
 RT group A Streptococcus strains associated with acute rheumatic fever  
 RT outbreaks.\*;

Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).  
 CC -1- SIMILARITY: BELONGS TO THE ERA/TRHE FAMILY OF GTP-BINDING  
 CC PROTEINS. ENGA SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-eb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE006498; AKK33393.1; -;  
 DR EMBL; AE009978; AL97088.1; -;  
 DR HANAP; MF00195; -1;  
 DR InterPro: IPR005289; GTP-binding\_dom.  
 DR InterPro: IPR006073; GTP\_OBG.  
 DR InterPro: IPR002917; MMR\_HSR1.  
 DR InterPro: IPR005225; Small-GTP.  
 DR Pfam; PF01926; MMR\_HSR1.1.  
 DR PRINTS; PR00326; GTP\_OBG.  
 DR TIGRfams; TIGR00650; MG442; 2.  
 DR TIGRfams; TIGR00231; small\_gtp; 2.  
 KW GTP-binding; Repeat; Complete proteome.  
 FT NP\_BIND 10 17 GTP 1 (POTENTIAL).  
 FT NP\_BIND 57 61 GTP 1 (POTENTIAL).  
 FT NP\_BIND 119 122 GTP 1 (POTENTIAL).  
 FT NP\_BIND 181 188 GTP 2 (POTENTIAL).  
 FT NP\_BIND 229 233 GTP 2 (POTENTIAL).  
 FT NP\_BIND 294 297 GTP 2 (POTENTIAL).  
 SO SEQUENCE 436 AA; 48801 MW; 59DEBD51675323A CRC64;

Query Match 69.0%; Score 1528; DB 1; Length 436;  
 Best Local Similarity 66.1%; Pred. No. 1,9e-86;  
 Matches 288; Conservative 75; Mismatches 71; Indels 2; Gaps 2;

1 MTKPIVAVGRPNVNGKSTIFNRIYGERVSYEDPTGRTORIRYSSGEMLTGHDNITDGG 60  
 1 MTKPIVAVGRPNVNGKSTIFNRIYGERVSYEDPTGRTORIRYSSGEMLTGHDNITDGG 60  
 61 IIGDAPFOTQIRAOAETIADEADYIIFPVNREGLTQSDENVAQILYKSKRPVLAANK 120  
 61 IIGDAPFOTQIRAOAETIADEADYIIFPVNREGLTQSDENVAQILYKSKRPVLAANK 120  
 61 IIGDAPFOTQIRAOAETIADEADYIIFPVNREGLTQSDENVAQILYKSKRPVLAANK 120  
 121 VDNMERTDVIDYDFYSLGFGEPYPIGSHGLGLDGLAVVSHFGEEDPDYEDTIRLSI 180  
 121 VDNMERTDVIDYDFYSLGFGEPYPIGSHGLGLDGLAVVSHFGEEDPDYEDTIRLSI 180  
 121 VDNMERTDVIDYDFYSLGFGEPYPIGSHGLGLDGLAVVSHFGEEDPDYEDTIRLSI 180  
 181 IGRPNVNGKSLVNAALGEDRVYVSNVAGTTPDAIDTGYSTODGDYVLTIDTACKRKKGY 240  
 181 IGRPNVNGKSLVNAALGEDRVYVSNVAGTTPDAIDTGYSTODGDYVLTIDTACKRKKGY 240  
 181 IGRPNVNGKSLVNAALGEDRVYVSNVAGTTPDAIDTGYSTODGDYVLTIDTACKRKKGY 240  
 240 YESTERYSVLRALKAIERSNVVLYDAEOGIIEDDKRVAGYAEHOGKAVVYVNMKDIVE 300  
 240 YESTERYSVLRALKAIERSNVVLYDAEOGIIEDDKRVAGYAEHOGKAVVYVNMKDIVE 300  
 240 YESTERYSVLRALKAIERSNVVLYDAEOGIIEDDKRVAGYAEHOGKAVVYVNMKDIVE 300  
 300 KDKTKMKFDEDEVRKPEFQIDYADAPASAKERTRLTLPYINAESENKRRVOSTLN 360  
 300 KDKTKMKFDEDEVRKPEFQIDYADAPASAKERTRLTLPYINAESENKRRVOSTLN 360  
 300 KDKTKMKFDEDEVRKPEFQIDYADAPASAKERTRLTLPYINAESENKRRVOSTLN 360  
 360 EYVTDATSNMPTPTDKGRRLNVYATOVAIEPTFVVVNDVLMHFSKRYLENOIRAA 419  
 360 EYVTDATSNMPTPTDKGRRLNVYATOVAIEPTFVVVNDVLMHFSKRYLENOIRAA 419  
 360 EYVTDATSNMPTPTDKGRRLNVYATOVAIEPTFVVVNDVLMHFSKRYLENOIRAA 419  
 420 AFGEETPIHIIIAKRN 435  
 420 AFGEETPIHIIIAKRN 435  
 420 AFGEETPIHIIIAKRN 435

RESULT 7  
 ENCA\_STRP3 STANDARD: PRT; 436 AA.

AC 08K878;  
 DT 28-FEB-2003 (rel. 41, Created)  
 DT 28-FEB-2003 (rel. 41, Last sequence update)  
 DT 28-FEB-2003 (rel. 41, Last annotation update)  
 DE Probable GTP-binding protein enga.  
 GN ENGA OR PGDA OR SPY0341 OR SPYM18\_0334.  
 OS Streptococcus pyogenes, and  
 OS Streptococcus pyogenes (serotype M18).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus;  
 ON NCBI\_TaxID=1314, 186103;  
 OX 111  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SF370 / ATCC 700294 / Serotype M1;  
 RC MEDLINE-21192684; PubMed-11296296;  
 RA Ferritelli J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
 RA Pliamex C., Szate S., Svorov A.N., Kenton S., Lai H.S., Lin S.P.,  
 RA Qian X., Clifton S.W., Roe B.A., McLaughlin R.,  
 RA Tian X., Clifton S.W., Roe B.A., McLaughlin R.,  
 RA Complete genome sequence of an M1 strain of Streptococcus pyogenes.\*;  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
 RN 121  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MGAS8233 / Serotype M18;  
 RC MEDLINE-21927593; PubMed-11917108;  
 RA Smoot J.C., Barblan K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
 RA Sylve G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,  
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,  
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.,  
 RT Genome sequence and comparative microarray analysis of serotype M18  
 RT group A Streptococcus strains associated with acute rheumatic fever  
 RT outbreaks.\*;

**CC** This SWISS-PROT entry is copyright © It is produced through a collaboration  
**CC**

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@slb-slb.ch](mailto:license@slb-slb.ch)).

CC EMBL: AE007464: AAK75787.1: -  
 CC EMBL: AE008533: AAL00357.1: -  
 CC PIR: B95199; B95199.  
 CC PIR: H98065; H98065.  
 CC TIGR: SP1709; -  
 CC HAMAP: MF\_00195; - 1.  
 CC InterPro: IPR005289; GTP-binding\_dom.  
 CC InterPro: IPR006073; GTP\_OBG.  
 CC InterPro: IPR002917; MMR\_HSR1.  
 CC InterPro: IPR005225; Small\_GTP.  
 CC Pfam: PF01926; MMR\_HSR1. 1.  
 CC PRINTS: PR00326; GTP\_OBG.  
 CC TIGRfams: TIGR00650; MG442; 2.  
 CC TIGRfams: TIGR00231; small\_GTP; 2.  
 CC GTP-binding; Repeat; Complete proteome.  
 CC NP\_BIND 10 GTP 1 (POTENTIAL).  
 CC NP\_BIND 57 61 GTP 1 (POTENTIAL).  
 CC NP\_BIND 119 122 GTP 1 (POTENTIAL).  
 CC NP\_BIND 181 188 GTP 2 (POTENTIAL).  
 CC NP\_BIND 229 233 GTP 2 (POTENTIAL).  
 CC NP\_BIND 294 297 GTP 2 (POTENTIAL).  
 CC NP\_BIND 297 297 GTP 2 (POTENTIAL).  
 CC SEQUENCE 436 AA: 49082 MW: 80138900528358BD CRC64:

Query Match 68.8% Score 1522; DB 1; Length 436;  
 Best Local Similarity 66.2%; Pred. No. 4.4e-86;

Matches 290; Conservative 73; Mismatches 69; Indels 6; Gaps 3;

OY 1 MTKPIVAIGRPVNGKSTIFNRIGERSIVEDPTGVTDRIRYSGEMLTNDPNIIDGG 60  
 DB 1 MALPTIAIGRPVNGKSTIFNRIGERSIVEDPTGVTDRIRYSGEMLTNDPNIIDGG 60  
 OY 61 IEIGDAPFOTOIRAOAETIADEAVIIFPNVNRGCLGSDENVAOILYKSKRPVLA 120  
 DB 61 IDVDAPFMEQIKHOAETIADEAVIIFPNVNRGCLGSDENVAOILYKSKRPVLA 120  
 OY 121 VDNEMKRDVDFYSLGEGEPYPISSGSHGLGDLDAVVSFSGE--DEEDPYEDRT 178  
 DB 121 VDNEMKRDVDFYSLGEGEPYPISSGSHGLGDLDAVVSFSGE--DEEDPYEDRT 178  
 OY 121 VDNEMKRDVDFYSLGEGEPYPISSGSHGLGDLDAVVSFSGE--DEEDPYEDRT 178  
 DB 121 VDNEMKRDVDFYSLGEGEPYPISSGSHGLGDLDAVVSFSGE--DEEDPYEDRT 178  
 OY 179 SIIGRPVNGKSSILNALIGEDRIYSGVAGTTRDAIDREYS-YDGDVILDTAGMRK 237  
 DB 178 SIIGRPVNGKSSILNALIGEDRIYSGVAGTTRDAIDREYS-YDGDVILDTAGMRK 237  
 OY 238 KYESTERYSVLRALAKAIERSNVVLVIADEOGIIEODKRVAGVAHEDGKAVV 297  
 DB 238 KYESTERYSVLRALAKAIERSNVVLVIADEOGIIEODKRVAGVAHEDGKAVV 297  
 OY 238 KYESTERYSVLRALAKAIERSNVVLVIADEOGIIEODKRVAGVAHEDGKAVV 297  
 DB 238 KYESTERYSVLRALAKAIERSNVVLVIADEOGIIEODKRVAGVAHEDGKAVV 297  
 OY 298 TVEKDSKTKMKFEDEVAKKEFOFLDYAOIAFVSAKERTRLTLFPYINEASENHRK 357  
 DB 298 TVEKDSKTKMKFEDEVAKKEFOFLDYAOIAFVSAKERTRLTLFPYINEASENHRK 357  
 OY 298 TVEKDSKTKMKFEDEVAKKEFOFLDYAOIAFVSAKERTRLTLFPYINEASENHRK 357  
 DB 298 TVEKDSKTKMKFEDEVAKKEFOFLDYAOIAFVSAKERTRLTLFPYINEASENHRK 357  
 OY 358 TLNEVYDAISMANPTFDKGRRLNVFATOVAILERTPEYVFNVDVLMHESKRL 417  
 DB 358 TLNEVYDAISMANPTFDKGRRLNVFATOVAILERTPEYVFNVDVLMHESKRL 417  
 OY 358 VLNQVDAISMANPTFDKGRRLNVFATOVAILERTPEYVFNVDVLMHESKRL 417  
 DB 358 VLNQVDAISMANPTFDKGRRLNVFATOVAILERTPEYVFNVDVLMHESKRL 417  
 OY 418 RAAFGFEGTPIHIIARKR 435  
 DB 418 RAAFGFEGTPIHIIARKR 435

RESULT 9  
 ENCA\_LACLA STANDARD: PRT: 436 AA.  
 AC 09CHH6:  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable GTP-binding protein enga.  
 ENCA OR YPHL OR LL0755.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria: Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 OX NCBI:TaxID=1360;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403;  
 RA MEDLINE=2135186; PubMed=11337471;  
 RA Bolotin A., Wincker P., Manger S., Jallou O., Malmgren K.,  
 RA Weissenbach J., Ehrlich S.D., Sorokin A.  
 RA "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RL lactis ssp. lactis IL1403."  
 CC Genome Res. 11:731-753(2001).  
 CC -1- SIMILARITY: BELONGS TO THE ERA/TRME FAMILY OF GTP-BINDING  
 CC PROTEINS. ENCA SUBFAMILY

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@slb-slb.ch](mailto:license@slb-slb.ch)).

CC EMBL: AE006309: AAK04853.1: -  
 CC PIR: C86719; C86719.  
 CC HAMAP: MF\_00195; - 1.  
 CC InterPro: IPR005289; GTP-binding\_dom.  
 CC InterPro: IPR006073; GTP\_OBG.  
 CC InterPro: IPR002917; MMR\_HSR1.  
 CC InterPro: IPR005225; Small\_GTP.  
 CC Pfam: PF01926; MMR\_HSR1. 1.  
 CC PRINTS: PR00326; GTP\_OBG.  
 CC TIGRfams: TIGR00650; MG442; 2.  
 CC TIGRfams: TIGR00231; small\_GTP; 2.  
 CC GTP-binding; Repeat; Complete proteome.  
 CC NP\_BIND 10 GTP 1 (POTENTIAL).  
 CC NP\_BIND 57 61 GTP 1 (POTENTIAL).  
 CC NP\_BIND 119 122 GTP 1 (POTENTIAL).  
 CC NP\_BIND 181 188 GTP 2 (POTENTIAL).  
 CC NP\_BIND 229 233 GTP 2 (POTENTIAL).  
 CC NP\_BIND 294 297 GTP 2 (POTENTIAL).  
 CC NP\_BIND 297 297 GTP 2 (POTENTIAL).  
 CC SEQUENCE 436 AA: 48866 MW: 1090081F966589D CRC64;

Query Match 68.7% Score 1520; DB 1; Length 436;  
 Best Local Similarity 66.3%; Pred. No. 5.9e-86;

Matches 289; Conservative 69; Mismatches 76; Indels 2; Gaps 2;

OY 1 MTKPIVAIGRPVNGKSTIFNRIGERSIVEDPTGVTDRIRYSGEMLTNDPNIIDGG 60  
 DB 1 MSLPTIAIGRPVNGKSTIFNRIGERSIVEDPTGVTDRIRYSGEMLTNDPNIIDGG 60  
 OY 61 IEIGDAPFOTOIRAOAETIADEAVIIFPNVNRGCLGSDENVAOILYKSKRPVLA 120  
 DB 61 IELSDPFTETIRAOAETIADEAVIIFPNVNRGCLGSDENVAOILYKSKRPVLA 120  
 OY 61 IELSDPFTETIRAOAETIADEAVIIFPNVNRGCLGSDENVAOILYKSKRPVLA 120  
 DB 61 IELSDPFTETIRAOAETIADEAVIIFPNVNRGCLGSDENVAOILYKSKRPVLA 120  
 OY 121 VDNEMKRDVDFYSLGEGEPYPISSGSHGLGDLDAVVSFSGE--DEEDPYEDRT 180  
 DB 121 VDNEMKRDVDFYSLGEGEPYPISSGSHGLGDLDAVVSFSGE--DEEDPYEDRT 180  
 OY 121 VDNEMKRDVDFYSLGEGEPYPISSGSHGLGDLDAVVSFSGE--DEEDPYEDRT 180  
 DB 121 VDNEMKRDVDFYSLGEGEPYPISSGSHGLGDLDAVVSFSGE--DEEDPYEDRT 180  
 OY 181 IGRPNVNGKSSILNALIGEDRIYSGVAGTTRDAIDREYS-YDGDVILDTAGMRK 239  
 DB 180 IGRPNVNGKSSILNALIGEDRIYSGVAGTTRDAIDREYS-YDGDVILDTAGMRK 239  
 OY 240 YESTERYSVLRALAKAIERSNVVLVIADEOGIIEODKRVAGVAHEDGKAVV 299  
 DB 240 YESTERYSVLRALAKAIERSNVVLVIADEOGIIEODKRVAGVAHEDGKAVV 299  
 OY 240 YESTERYSVLRALAKAIERSNVVLVIADEOGIIEODKRVAGVAHEDGKAVV 299  
 DB 240 YESTERYSVLRALAKAIERSNVVLVIADEOGIIEODKRVAGVAHEDGKAVV 299  
 OY 300 EKDSKTKMKFEDEVAKKEFOFLDYAOIAFVSAKERTRLTLFPYINEASENHRK 359  
 DB 300 EKDSKTKMKFEDEVAKKEFOFLDYAOIAFVSAKERTRLTLFPYINEASENHRK 359

OY 360 NEVYDAISMNPTPTDGRRLNRYATOVAIPEPTFVFNVDVLMHFSYKRLNIOIRA 419  
DB 360 NDVIMDAVINPTPTDKGRRLKIFATOVAIKPTTFVFNVEELMHFSYKRLNIOIRA 419  
OY 420 AFCECTPIHIIARRK 435  
DB 420 AFVECTPVHLIARRK 435

## RESULT 10

ENCA\_THETN STANDARD: PRT: 439 AA.

AC OR9J1;  
DT 28-FEB-2003 (rel. 41, Created)  
DT 28-FEB-2003 (rel. 41, Last sequence update)  
DE 28-FEB-2003 (rel. 41, Last annotation update)  
DE Probable GTP-binding protein engA.  
GN ENCA OR TTE1619.  
OS Thermoaerobacter tengcongensis.  
OC Bacteria: Firmicutes: Clostridia: Thermoaerobacteriales;  
OC Thermoaerobacteriaceae; Thermoaerobacter.  
NCBI\_TaxID=119072;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=MB4 / JCM 11007;  
RX MEDLINE=21992816; PubMed=11997336;  
RA Bao Q., Tian Y., Li W., Xu Z., Hu S., Dong W., Yang J.,  
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
RA Ten H., Chen R., Wang J., Yu J., Yang H.;  
RT "A complete sequence of T. tengcongensis genome.";  
RL Genome Res. 13:689-700(2002)  
CC -1- SIMILARITY: BELONGS TO THE ERA/TRME FAMILY OF GTP-BINDING  
CC PROTEINS. ENCA SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: AEO13117; AAM24821.1; -  
DR HAMAP: MF\_00195; -; 1.  
DR InterPro: IPR003593; AAA\_ATPase.  
DR InterPro: IPR003289; GTP-binding\_dom.  
DR InterPro: IPR002917; MMR\_HSR1.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF01926; MMR\_HSR1. 1.  
DR SMART: SM00382; AAA; 2.  
DR TIGRfams: TIGR00650; MG442; 2.  
DR TIGRfams: TIGR00231; small\_GTP; 2.  
KM GTP-binding: Repeat: Complete proteome.  
FT NP\_BIND 10 17 GTP 1 (POTENTIAL).  
FT NP\_BIND 57 61 GTP 1 (POTENTIAL).  
FT NP\_BIND 120 123 GTP 1 (POTENTIAL).  
FT NP\_BIND 183 190 GTP 2 (POTENTIAL).  
FT NP\_BIND 230 234 GTP 2 (POTENTIAL).  
FT NP\_BIND 295 298 GTP 2 (POTENTIAL).  
SQ SEQUENCE 439 AA: 49432 MW: 77052 FOF268DC5 CRC64:

Query Match 56.6% Score 1251.5; DR 1; Length 439;  
Best Local Similarity 54.1% Pred. No. 1.5e-69;  
Matches 237; Conservative 92; Mismatches 104; Indels 5; Gaps 4;

OY 1 MTKPVAIVGRPNVGSSTFNRIVGERSVIEDTPEGTRDRIVSSGEMLTDFHNIIDTGC 60  
DB 1 MSRAVVAIVGRPNVGSSTFNRIVGERSVIEDTPEGTRDRIVGNAEMLRKFFILVDTCG 60  
OY 61 IEIG--DAPFOTIRAOAEIADAVIIFMVNVNREGTQSDMVAQIILYKSKKPVAVL 118  
DB 61 LDPPKQDVIF-SKVYKLVGEAIEAADIILFVVDAREGLVPEDEEIANMLRKTKKEVILVC 119

OY 119 NKVDNM-EMKTDVYDFYSLGCEGPYPISGSHGLGDLDAVNSHFCEEDPYDEDTIR 177  
DB 120 NKVDSFKEMKASFYDFEKLGEPIPIISASNGLGIGELLDVIRLENVDE-YEEETIV 178  
OY 178 LSIIGRPNVGSSLVNVAIIGEDRIVSNVAGTTRADITREXSYGQDVLIDTGNKRG 237  
DB 179 IAVIGRPNVGSSLVNVAIIGEDRIVSDIPIGTTRADITPTKGRNIIIDTGNKRG 238  
OY 238 KYVSTKYSVLALAKAIERSNVVVIDAEGIIIDOKRAGTAHEGKAAVIVVAKMD 297  
DB 239 RISESIERYSVLALAIERADICLMDATGTEGDTKJAGAFENGIIILVAKMD 298  
OY 298 IVEKDSNTYKEYTEMIREKIAFISFAPILFISAKTGRIHVLFTVOKVMEVNRKITTG 358  
DB 358 TINEVYVDASMNPTPTDGRRLNRYATOVAIPEPTFVFNVDVLMHFSYKRLNIOIRA 417  
DB 359 LNNVLEAMLMPPSPSSGRPKITVATOVGTAKPTFVFNVEELMHFSYKRLNIOIRA 418  
OY 418 RAAFCECTPIHIIARRK 435  
DB 419 RONGCECTPIVISTKRK 436

## RESULT 11

ENCA\_CLOPE STANDARD: PRT: 438 AA.

AC OR9J1;  
DT 28-FEB-2003 (rel. 41, Created)  
DT 28-FEB-2003 (rel. 41, Last sequence update)  
DE 28-FEB-2003 (rel. 41, Last annotation update)  
DE Probable GTP-binding protein engA.  
GN ENCA OR CPE1755.  
OS Clostridium perfringens.  
OC Bacteria: Firmicutes: Clostridia: Clostridiales; Clostridiaceae;  
OC Clostridium.  
NCBI\_TaxID=1502;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=13 / Type A;  
RX MEDLINE=21664373; PubMed=11792842;  
RA Shimizu T., Onoani K., Hiraoka H., Ohshima K., Yamashita A.,  
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
RT flesh-eater.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
CC -1- SIMILARITY: BELONGS TO THE ERA/TRME FAMILY OF GTP-BINDING  
CC PROTEINS. ENCA SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: AP003191; BAB81461.1; -  
DR HAMAP: MF\_00195; -; 1.  
DR InterPro: IPR003593; AAA\_ATPase.  
DR InterPro: IPR005289; GTP-binding\_dom.  
DR InterPro: IPR006073; GTP\_OBG.  
DR InterPro: IPR002917; MMR\_HSR1.  
DR InterPro: IPR001806; Raa\_crsfmg.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF01926; MMR\_HSR1. 1.  
DR PRINTS: PR00326; GTP\_OBG.  
DR PRINTS: PR00449; RASTRNSFRMNC.  
DR SMART: SM00382; AAA; 2.  
DR TIGRfams: TIGR00650; MG442; 2.  
DR TIGRfams: TIGR00231; small\_GTP; 2.





28-FEB-2003 (Rel. 41, last annotation update)  
 DE Probable GTP-binding protein enga.  
 GN ENGA OR FN0170.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 CC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;  
 CC Fusobacterium.  
 CC NCBI\_TaxID=76856;  
 RN 121  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 25586.  
 RX MEDLINE-21886394; PubMed-11889109.  
 RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lygidis A.,  
 RA Bhattacharya A., Bartman A., Gardner W., Gueckin G., Zhu L.,  
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,  
 RA Larsen N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,  
 RA Fongstein M., Kyrides N., Overbeek R.,  
 RA "Genome sequence and analysis of the oral bacterium Fusobacterium  
 RT nucleatum strain ATCC 25586."  
 RL J. Bacteriol. 184:2005-2018(2002).  
 CC -1 SIMILARITY: BELONGS TO THE ERA/TIME FAMILY OF GTP-BINDING  
 CC PROTEINS. ENGA SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR HAMAP: MF\_00195; AAL94376.1; -  
 DR InterPro: IPR005289; GTP-binding\_dom.  
 DR InterPro: IPR002917; MMR\_HSR1.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF01926; MMR\_HSR1; 1.  
 DR TIGRFAMs: TIGR00650; MG442; 2.  
 DR TIGRFAMs: TIGR00231; small\_GTP; 2.  
 KW GTP-binding; Repeat; Complete proteome.  
 FT NP\_BIND 9 16 GTP 1 (POTENTIAL).  
 FT NP\_BIND 56 60 GTP 1 (POTENTIAL).  
 FT NP\_BIND 119 132 GTP 1 (POTENTIAL).  
 FT NP\_BIND 183 190 GTP 2 (POTENTIAL).  
 FT NP\_BIND 230 234 GTP 2 (POTENTIAL).  
 FT NP\_BIND 295 298 GTP 2 (POTENTIAL).  
 FT SEQUENCE 440 AA: 49608 MW: 86205f839b2a05 CRC64;  
 SO  
 Query Match 53.3%; Score 1178.5; DB 1; Length 440;  
 Best Local Similarity 53.6%; Pred. No. 4:3e-65;  
 Matches 231; Conservative 81; Mismatches 116; Indels 3; Gaps 3;  
 Oy 3 KPIVAIVGRPNVGRKSTFNRIVGRSVIEDTQGVTRDRIRYSSEGLMTHDFNIIIDGGIE 62  
 Db 2 KPIAIVGRPNVGRKSTFNRIVGRSVIEDTQGVTRDRIRYSGEGLMTHDFNIIIDGGIE 61  
 Oy 63 IGDAPV-QTOIRAOEIAIDADVIIIPVNRREGITQSDENVAQIILYKSKPPVLAIVNKY 121  
 Db 62 PRNNDLMAKIKEQAEVANNENADVILFVVDKSGNPNPDEIAIYLKKNKPPVILCVNKI 121  
 Oy 122 DN-MEARTDVDFYSIGFGEPRYSGHGLGDLIDLVVHFEEEDPDDEDTIRLST 180  
 Db 122 DNFEQDDVDYDFYIGFGEPRYSGHGLGDLIDLVVHFEEEDPDDEDTIRLST 181  
 Oy 181 IGRPNVGRKSSLYNALIGEDRVIVNACTTDAIDTFEYSDGQDYVLIDTAKMKKKYV 240  
 Db 182 IGRPNVGRKSSLYNALIGEDRVIVNACTTDAIDTFEYSDGQDYVLIDTAKMKKKYV 241  
 Oy 241 ESTEKYSVLRALKATERSNVVLVDAEGITTEDKRVAGYAHGKAAVYIVNKKDQVE 300  
 Db 242 ESTEYYSVLRALKATERSNVVLVDAEGITTEDKRVAGYAHGKAAVYIVNKKDQVE 301  
 Oy 301 -KDSKTKMKFEDEYKREOFLDYAOIAYVSAKERTRLTFTPIYENASNNKKRVOSSTL 359  
 Db 301 -KDSKTKMKFEDEYKREOFLDYAOIAYVSAKERTRLTFTPIYENASNNKKRVOSSTL 359

Db 302 NKNNAKTKKKEELVLEPLSLAPLEFVSALTGRTNLLLEIDRIYEETKRISTGLL 361  
 Oy 360 NEVYTDALSMNPPTDQGRRLNRYATQVALEPPFVFNVDVLEHMFYKRLNQIRA 419  
 Db 362 NTLIKDVAVLNNPPTKGRVIXINATQVSPAPFVLCNYPPELLHFSYARVYENKFR 421  
 Oy 420 AFGEGPTPIHI 430  
 Db 422 AFGDGSPIHI 432  
 RESULT 14  
 ENGA\_ANASP STANDARD; PRT: 453 AA.  
 ID ENGA\_ANASP  
 AC Q8YEH7;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, last sequence update)  
 DE 28-FEB-2003 (Rel. 41, last annotation update)  
 DE Probable GTP-binding protein enga.  
 GN ENGA OR ALR0483  
 OS Anabaena sp. (strain PCC 7120).  
 CC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 CC NCBI\_TaxID=103690;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-21595285; PubMed-11759840;  
 RA Keneke T., Nakamura Y., Wolk C.P., Kawashima K., Kimura T.,  
 RA Watanabe A., Iriguchi M., Ishikawa A., Matsuno M., Muroki A.,  
 RA Kishida Y., Kohara S., Sugimoto M., Takazawa M., Yanada M.,  
 RA Yasuda M., Tabata S.,  
 RA "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium Anabaena sp. strain PCC 7120."  
 RL DNA Res. 8:205-213(2001).  
 CC -1 SIMILARITY: BELONGS TO THE ERA/TIME FAMILY OF GTP-BINDING  
 CC PROTEINS. ENGA SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AP003582; BAB72441.1; -  
 DR HAMAP: MF\_00195; AAL94376.1; -  
 DR InterPro: IPR005289; GTP-binding\_dom.  
 DR InterPro: IPR006073; GTP1\_OBG.  
 DR InterPro: IPR002917; MMR\_HSR1.  
 DR InterPro: IPR001806; Ras\_transfmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF01926; MMR\_HSR1; 1.  
 DR PRINTS: PR00326; GTP1\_OBG.  
 DR TIGRFAMs: TIGR00650; MG442; 2.  
 DR TIGRFAMs: TIGR00231; small\_GTP; 2.  
 KW GTP-binding; Repeat; Complete proteome.  
 FT NP\_BIND 10 17 GTP 1 (POTENTIAL).  
 FT NP\_BIND 57 61 GTP 1 (POTENTIAL).  
 FT NP\_BIND 120 123 GTP 1 (POTENTIAL).  
 FT NP\_BIND 183 190 GTP 2 (POTENTIAL).  
 FT NP\_BIND 230 234 GTP 2 (POTENTIAL).  
 FT NP\_BIND 295 298 GTP 2 (POTENTIAL).  
 FT SEQUENCE 453 AA: 50731 MW: 23432f10c2e287d1 CRC64;  
 SO  
 Query Match 46.9%; Score 1037; DB 1; Length 453;  
 Best Local Similarity 47.6%; Pred. No. 2e-56;  
 Matches 210; Conservative 83; Mismatches 138; Indels 10; Gaps 5;  
 Oy 1 MTRPIAIVGRPNVGRKSTFNRIVGRSVIEDTQGVTRDRIRYSSEGLMTHDFNIIIDGG 60  
 Db 1 MTRPIAIVGRPNVGRKSTFNRIVGRSVIEDTQGVTRDRIRYSSEGLMTHDFNIIIDGG 60



```

Db      1 MGLEPIAIIIGRPVNGKSTLVNRLAGEQTAIVHDEPGVTRDRTTYLPAYNSDBEFQVDTGC 60
Oy      61 IEIG-DAPFOTOIARAQAEIAIDADVIFPVNVRREGLTQSDENVAQIILYKSKRPVLAVN 119
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      61 LVFNDDREFLPLIRQOALAAHBSAAIFVNGQGPNSADEIAEMLRQOPVPEVLAVN 120
Oy      120 KVDNMEMRT-DVYDFSLGCEPYPISGSHGLGDLDAVAVSHF---GSEEDPYDEDT 175
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      121 KCSEPDGSGIOASEFPELGLGEYFPLSAIHNGSTGKILPPTLEEN----NE 176
Oy      176 IRLSIIGRPVNGKSSLVNALGEDRVYVSNVAGTTRDAIDTEYSYDGDVYLIDTAGNRK 235
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      177 IKAIIGRPVNGKSSLVNALGAGEERYVISPISGTRDAIDTEIEEDGCONYLIDTAGIRK 236
Oy      236 KGKVESTKYSVLRALKAIERSNVLYVIDAOGIIEODKRVAGAHGOKAVYIVNKK 295
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      237 KKSIDYGTETEPFSINRAFKAIRADVYLVIDALDGYTEDODKLAGRIIDEGKACVYVYNNK 296
Oy      296 WDVEKDSKTKMKFEDEVRKEKRFQFLDYAOIAFVSAKERTRLTFLPYINEASENKKRVQ 355
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      297 WDAVERKDSYTIYDEKLEARLHFTEMADTIYVSAVTGQVREKILELVTKANEBKRRVS 356
Oy      356 SSTLNEYVTDAISNPTPTDK-GRLLNFFATQVAIEPTTFVFNVDYELMHFSYKRYLE 414
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      357 TSVINEYLEDVAVMHSPPTRSGRGGRITTYGTGYSTOPTALFVNEAKRRNDYRRITIE 416
Oy      415 NQIRAAFGCEGTPIHIAKR 435
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      417 RFRKQOLGFGKPTIRLLMSK 437

```

```

RESULT 15
ENGA_SVNV3 STANDARD: PRT: 452 AA.
AC P74120:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable GTP-binding protein enga.
GN ENGA OR SLR1974.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97061201: Pubmed-8905231:
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shampo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT *Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.*;
RT DNA Res. 3:109-136(1996).
CC -1 SIMILARITY: BELONGS TO THE ERA/TIME FAMILY OF GTP-BINDING
CC PROTEINS. ENGA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90912; BAA18206.1;
CC PIR: S75645; S75645.
CC HAMAP: MF_00195; 1.
CC InterPro: IPR003593; AAA_ATPase.
CC InterPro: IPR003289; GTP-binding_dom.
CC InterPro: IPR006073; GTP_OBG.
CC InterPro: IPR002917; MMR_HSR1.

```

```

DR InterPro: IPR001806; Ras_trnsmfmg.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF01926; MMR_HSR1.
DR PRINTS: PR00326; GTP_OBG.
DR SMART: SM00382; AAA_2.
DR SMART: SM00382; RASTRNSFRMNG.
DR TIGRfams: TIGR00650; M6442_2.
DR TIGRfams: TIGR00231; small_gtp_2.
KW GTP-binding; Repeat; Complete proteome.
FT NP_BIND 10 17 GTP 1 (POTENTIAL).
FT NP_BIND 57 61 GTP 1 (POTENTIAL).
FT NP_BIND 120 123 GTP 1 (POTENTIAL).
FT NP_BIND 183 190 GTP 2 (POTENTIAL).
FT NP_BIND 230 234 GTP 2 (POTENTIAL).
FT NP_BIND 295 298 GTP 2 (POTENTIAL).
SO SEQUENCE 452 AA; 50825 MW; 7FF711EAF7D6CC76 CRC64;

```

```

Query Match 46.18; Score 1021; DB 1; Length 452;
Best Local Similarity 45.98; Pred. No. 1,9e-55;
Matches 200; Conservative 94; Mismatches 138; Indels 4; Gaps 4;

```

```

Oy      1 MTKPIVAIGRPVNGKSTLVNRLAGEQTAIVHDEPGVTRDRTTYLPAYNSDBEFQVDTGC 60
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1 MSLPIVAILIGRPVNGKSTLVNRLAGEQTAIVHDEPGVTRDRTTYLPAYNSDBEFQVDTGC 60
Oy      61 IEIG-DAPFOTOIARAQAEIAIDADVIFPVNVRREGLTQSDENVAQIILYKSKRPVLAVN 119
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      61 LVFNDDREFLPLIRQOALAAHBSAAIFVNGQGPNSADEIAEMLRQOPVPEVLAVN 120
Oy      120 KVDNMEMRT-DVYDFSLGCEPYPISGSHGLGDLDAVAVSHF---GSEEDPYDEDT 175
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      121 KCSEPDGSGIOASEFPELGLGEYFPLSAIHNGSTGKILPPTLEEN----NE 176
Oy      176 IRLSIIGRPVNGKSSLVNALGEDRVYVSNVAGTTRDAIDTEYSYDGDVYLIDTAGNRK 235
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      177 IKAIIGRPVNGKSSLVNALGAGEERYVISPISGTRDAIDTEIEEDGCONYLIDTAGIRK 236
Oy      236 KGKVESTKYSVLRALKAIERSNVLYVIDAOGIIEODKRVAGAHGOKAVYIVNKK 295
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      237 KKSIDYGTETEPFSINRAFKAIRADVYLVIDALDGYTEDODKLAGRIIDEGKACVYVYNNK 296
Oy      296 WDVEKDSKTKMKFEDEVRKEKRFQFLDYAOIAFVSAKERTRLTFLPYINEASENKKRVQ 355
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      297 WDAVERKDSYTIYDEKLEARLHFTEMADTIYVSAVTGQVREKILELVTKANEBKRRVS 356
Oy      356 SSTLNEYVTDAISNPTPTDK-GRLLNFFATQVAIEPTTFVFNVDYELMHFSYKRYLE 414
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      357 TSVINEYLEDVAVMHSPPTRSGRGGRITTYGTGYSTOPTALFVNEAKRRNDYRRITIE 416
Oy      415 NQIRAAFGCEGTPIHIAKR 435
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      417 RFRKQOLGFGKPTIRLLMSK 437

```

Search completed: September 10, 2003, 00:33:14  
Job time : 25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 00:29:31; Search time 100 Seconds

(without alignments)  
1125.109 Million cell updates/sec

Title: US-09-815-242-12600

Perfect score: 2213

Sequence: 1 MTKPIVAIVGRPNVGRKSTIF.....IRAPGFGPTPIHIAKRN 436

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database: 1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MNC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PROTOZOA:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEA:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2094	94.6	436	16	08CP62
2	1614	72.9	436	16	08EQAB
3	1523	68.8	436	16	08E3T9
4	1523	68.8	436	16	08DY73
5	1476	66.7	436	2	09RHV5
6	1472	66.5	436	16	08DS90
7	1342	60.6	445	2	08GES8
8	1340.5	60.6	435	2	08KH12
9	1028.5	46.5	449	16	08DK11
10	1022	46.2	444	16	08EPM6
11	920	41.6	503	16	08EFS9
12	898.5	40.6	496	16	08DF02
13	874	39.5	537	10	09LH58
14	874	39.5	659	10	09C7C0
15	859	38.8	487	16	08EC36
16	851.5	38.5	463	16	08G6A8

17	843.5	38.1	552	16	08EPM5	08EPM5 corynebacte
18	838	37.9	588	16	09A7R6	09A7R6 caulobacter
19	796.5	36.0	483	16	08G2E8	08G2E8 brucella su
20	756.5	34.2	874	5	08ISN5	08ISN5 plasmodium
21	736	33.3	496	10	09AW74	09AW74 guillierdia
22	678	30.6	489	16	08E6K1	08E6K1 leptospira
23	661	29.9	188	11	09E212	09E212 lactospora
24	559.5	25.3	383	2	P94645	P94645 raltus morv
25	524	23.7	642	10	09AX00	09AX00 oryza sativ
26	514.5	23.2	514	10	09ELED	09ELED arabidopsis
27	512	23.1	347	2	08VN97	08VN97 helicobacte
28	512	23.1	347	2	08VN36	08VN36 helicobacte
29	511	23.1	347	2	08VN38	08VN38 helicobacte
30	455	20.6	281	2	09LA73	09LA73 thiolobact
31	435	19.7	456	16	08D1Y0	08D1Y0 violognesw
32	433.5	19.6	208	2	P72518	P72518 synechococ
33	320.5	14.5	190	2	09AEJ8	09AEJ8 campylobact
34	320.5	14.5	190	2	09AQ16	09AQ16 campylobact
35	319.5	14.4	190	2	09AEJ6	09AEJ6 campylobact
36	318.5	14.4	190	2	09AEJ9	09AEJ9 campylobact
37	316.5	14.3	190	2	09AEJ7	09AEJ7 campylobact
38	266	12.0	479	16	08DPZ8	08DPZ8 streptococ
39	258	11.7	300	16	08EPY0	08EPY0 oceanobact
40	243.5	11.0	455	16	08DPT8	08DPT8 streptococ
41	241.5	10.9	170	2	09S3C9	09S3C9 helicobacte
42	241.5	10.9	170	2	09S3C8	09S3C8 helicobacte
43	241.5	10.9	170	2	09S3C7	09S3C7 helicobacte
44	241.5	10.9	170	2	08VNA0	08VNA0 helicobacte
45	241.5	10.9	170	2	09R3H9	09R3H9 helicobacte

## ALIGNMENTS

## RESULT 1

08CP62 ID 08CP62 PRELIMINARY: PRT: 436 AA.  
AC 08CP62;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE GTP binding protein.  
GN SE1163.  
OS Streptococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Streptococcus.  
OX NCBI\_TaxId=1282;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Chen Z., Ren Y., Li H., Fu G., Lu L., Lu G., Jia J., Yu Y., Qin Z.,  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE016747; AAC04760.1; -  
KW Complete proteome.  
SQ SEQUENCE 436 AA: 49024 MM: 77C74B95641D3F22 CRC64:

Query Match 94.64; Score 2094; DB 16; Length 436;  
Best Local Similarity 93.18; Pred. No. 2.9e-124;  
Matches 406; Conservative 18; Mismatches 12; Indels 0; Gaps 0;  
QY 1 MTKPIVAIVGRPNVGRKSTIFNRIVGERVSIEDPCVTRDRYSSGEMLTDFNIDTGC 60  
|||||  
1 MTKPIVAIVGRPNVGRKSTIFNRIVGERVSIEDPCVTRDRYSSGEMLTDFNIDTGC 60  
Db 1 MTKPIVAIVGRPNVGRKSTIFNRIVGERVSIEDPCVTRDRYSSGEMLTDFNIDTGC 60  
QY 61 IRTGAPFOTIRAOEATIDEAVITFENVNREGLTQSDENVAOILYKSKPPVLAANK 120  
|||||  
61 IRTGAPFOTIRAOEATIDEAVITFENVNREGLTQSDENVAOILYKSKPPVLAANK 120  
Db 61 IRTGAPFOTIRAOEATIDEAVITFENVNREGLTQSDENVAOILYKSKPPVLAANK 120  
QY 121 VDNMEKRDYVDYFSLGEGEPYPISSHGGLGDLDDAVVSHGEEEDPYDDETIRLSI 180  
|||||  
121 VDNMEKRDYVDYFSLGEGEPYPISSHGGLGDLDDAVVSHGEEEDPYDDETIRLSI 180  
Db 121 VDNMEKRDYVDYFSLGEGEPYPISSHGGLGDLDDAVVSHGEEEDPYDDETIRLSI 180  
QY 181 IGRPNVGRKSLVNMILGSDRYVSNVAGTTRDAIDREYSYDQDQVLLIDTAGMRKKGVY 240

DB 181 ICRPNVGSLLVNALGGERIVSNAGCTTRDAIDTETSTGDDVYLIDTAGKRRKKGY 240  
OY 241 ESTERYSVLRALKAIERSNVVLYIDAEGCIIEODKRVAGYAHBOGKAVIYVNMKDYVE 300  
DB 241 ESTERYSVLRALKAIERSNVVLYIDAEGCIIEODKRVAGYAHBOGKAVIYVNMKDYVE 300  
OY 301 KDSKTKMKFFEDVREKREFQFLDYAOIAFVSAKERTRLTLPFYINEASENHKKRVOSSTLN 360  
DB 301 KDSKTKMKFFEDVREKREFQFLDYAOIAFVSAKERTRLTLPFYINEASENHKKRVOSSTLN 360  
OY 361 EYVTDALISNMPPTDCKGRNLNVFATOVATIEPTTFVYVNDVELMHFSTKRYLENDIRAA 420  
DB 361 EYVTDALISNMPPTDCKGRNLNVFATOVATIEPTTFVYVNDVELMHFSTKRYLENDIRAA 420  
OY 421 FGEFGTPIHIIARKRN 436  
DB 421 FGEFGTPIHIIARKRN 436

## RESULT 2

DB 08E0A8 PRELIMINARY: PRT: 436 AA.  
AC 08E0A8: 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
GN O81797.  
OS Oceanobacillus theysensis.  
OC Bacteria: Firmicutes: Bacillales: Oceanobacillus.  
OX NCBI\_Taxid=182710.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HTE831 / DSM 14371 / JCM 11309;  
RX MEDLINE-22220767; PubMed-12235376;  
RA Takami H., Takaki Y., Uchiyama I.;  
RT "Genome sequence of Oceanobacillus theysensis isolated from the Iheya  
RT Ridge and its unexpected adaptive capabilities to extreme  
RT environments.";  
RL Nucleic Acids Res. 30:3927-3935 (2002).  
DR EMBL: AP004599; BAC13753.1; -  
KW Complete proteome.  
SQ SEQUENCE 436 AA: 48998 MW: 18F52E40D9B0C9C2 CRC64:

Query Match 72.9% Score 1614; DB 16; Length 436;  
Best Local Similarity 69.4%; Pred. No. 5.8e-94;  
Matches 302; Conservative 66; Mismatches 67; Indels 0; Gaps 0;

OY 1 MTKPIVIAIVGRPNVGSSTIFNRIIGERVSIVEDTPGVTRORIYSSGEMLTNDPNIIDTGC 60  
DB 1 MTKPIVIAIVGRPNVGSSTIFNRIIGERVSIVEDTPGVTRORIYSSGEMLTNDPNIIDTGC 60  
OY 61 IEGDAPFOTOIRAOAELAIIDEADVIIFMNVNREGCTOSDEMVAOILYKSKRPVLAANK 120  
DB 61 IEGDAPFOTOIRAOAELAIIDEADVIIFMNVNREGCTOSDEMVAOILYKSKRPVLAANK 120  
OY 121 VDNMKRRTVYDFYSLGCEPYPISGSHGLGLDLDLDAVYSHGGEEDPYPYEDTIRLSI 180  
DB 121 VDNMKRRTVYDFYSLGCEPYPISGSHGLGLDLDLDAVYSHGGEEDPYPYEDTIRLSI 180  
OY 181 IGRPNVGSLLVNALGGERIVSNAGCTTRDAIDTETSTGDDVYLIDTAGKRRKKGY 240  
DB 181 IGRPNVGSLLVNALGGERIVSNAGCTTRDAIDTETSTGDDVYLIDTAGKRRKKGY 240  
OY 241 ESTERYSVLRALKAIERSNVVLYIDAEGCIIEODKRVAGYAHBOGKAVIYVNMKDYVE 300  
DB 241 ESTERYSVLRALKAIERSNVVLYIDAEGCIIEODKRVAGYAHBOGKAVIYVNMKDYVE 300  
OY 301 KDSKTKMKFFEDVREKREFQFLDYAOIAFVSAKERTRLTLPFYINEASENHKKRVOSSTLN 360  
DB 301 KDSKTKMKFFEDVREKREFQFLDYAOIAFVSAKERTRLTLPFYINEASENHKKRVOSSTLN 360

OY 361 EYVTDALISNMPPTDCKGRNLNVFATOVATIEPTTFVYVNDVELMHFSTKRYLENDIRAA 420  
DB 361 EYVTDALISNMPPTDCKGRNLNVFATOVATIEPTTFVYVNDVELMHFSTKRYLENDIRAA 420  
OY 421 FGEFGTPIHIIARKRN 435  
DB 421 FGEFGTPIHIIARKRN 435

## RESULT 3

DB 08E379 PRELIMINARY: PRT: 436 AA.  
AC 08E379: 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
GN G851667.  
OS Streptococcus agalactiae (serotype III).  
OC Bacteria: Firmicutes: Lactobacillales: Streptococcaceae;  
OX Streptococcus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NEM316 / Serotype III;  
RX MEDLINE-2242508; PubMed-12354221;  
RA Glaser P., Rusnlok C., Buchrieser C., Chevalier F., Frangeul L.,  
RA Masdek T., Zouine M., Couve E., Lallou L., Poyart C., Tillet C.,  
RA Kunst F.;  
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
RT invasive neonatal disease.";  
RL MOL. Microbiol. 45:1499-1513 (2002).  
DR EMBL: AL766852; CAD47326.1; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 436 AA: 48880 MW: E7E333B4D0F3F610 CRC64:

Query Match 68.8% Score 1523; DB 16; Length 436;  
Best Local Similarity 66.5%; Pred. No. 3.2e-88;  
Matches 290; Conservative 70; Mismatches 74; Indels 2; Gaps 2;

OY 1 MTKPIVIAIVGRPNVGSSTIFNRIIGERVSIVEDTPGVTRORIYSSGEMLTNDPNIIDTGC 60  
DB 1 MTKPIVIAIVGRPNVGSSTIFNRIIGERVSIVEDTPGVTRORIYSSGEMLTNDPNIIDTGC 60  
OY 61 IEGDAPFOTOIRAOAELAIIDEADVIIFMNVNREGCTOSDEMVAOILYKSKRPVLAANK 120  
DB 61 IEGDAPFOTOIRAOAELAIIDEADVIIFMNVNREGCTOSDEMVAOILYKSKRPVLAANK 120  
OY 121 VDNMKRRTVYDFYSLGCEPYPISGSHGLGLDLDLDAVYSHGGEEDPYPYEDTIRLSI 180  
DB 121 VDNMKRRTVYDFYSLGCEPYPISGSHGLGLDLDLDAVYSHGGEEDPYPYEDTIRLSI 180  
OY 181 IGRPNVGSLLVNALGGERIVSNAGCTTRDAIDTETSTGDDVYLIDTAGKRRKKGY 240  
DB 181 IGRPNVGSLLVNALGGERIVSNAGCTTRDAIDTETSTGDDVYLIDTAGKRRKKGY 240  
OY 241 ESTERYSVLRALKAIERSNVVLYIDAEGCIIEODKRVAGYAHBOGKAVIYVNMKDYVE 300  
DB 241 ESTERYSVLRALKAIERSNVVLYIDAEGCIIEODKRVAGYAHBOGKAVIYVNMKDYVE 300  
OY 301 KDSKTKMKFFEDVREKREFQFLDYAOIAFVSAKERTRLTLPFYINEASENHKKRVOSSTLN 360  
DB 301 KDSKTKMKFFEDVREKREFQFLDYAOIAFVSAKERTRLTLPFYINEASENHKKRVOSSTLN 360  
OY 361 EYVTDALISNMPPTDCKGRNLNVFATOVATIEPTTFVYVNDVELMHFSTKRYLENDIRAA 420  
DB 361 EYVTDALISNMPPTDCKGRNLNVFATOVATIEPTTFVYVNDVELMHFSTKRYLENDIRAA 420

```

RESULT 4
08DY73
ID 08DY73 PRELIMINARY: PRT: 436 AA.
AC 08DY73
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Phosphoglycerate dehydrogenase-related protein.
GN SAG1620.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Maignan V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Messers L.C., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beaman M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Redune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carey H.A., Cline R.T., Van Allen S.E., Gill J., Scarselli M., Mora M.,
RA Jacobini E.T., Bretonni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo D., Rappunli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae,"
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL: AE014265; ANND0484.1; -.
DR TIGR: SAG1620; -.
KW Complete proteome.
KW SEQUENCE 436 AA; 48981 MW; E2C064B3F04B8644 CRC64;

```

Query Match	68.80;	Score 1523;	DB 16;	Length 436;
Best Local Similarity	66.30;	Pred. No. 3.2e-88;		
Matches 289;	Conservative 72;	Mismatches 73;	Indels 2;	Gaps 2

[illegible]

## RESULT 5

```

O9RHV5      O9RHV5      PRELIMINARY;      PRT;      436 AA.
AC      O9RHV5;
DT      01-MAY-2000 (TrEMBLrel. 13, Created)
DT      01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT      01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE      Phosphoglycerate dehydrogenase.
GN
GS
OS      Streptococcus mutans.
OC      Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC      Streptococcus.
OX      NCBI_TaxID=1309;
RN      11
RP      SEQUENCE FROM N.A.
RC      STRAIN-M78148;
RA      Kawabata S., Terao Y., Hamada S.;
RT      Molecular cloning, sequence and characterization of a novel
RL      streptococcal phosphoglycerate dehydrogenase gene.";
RL      EMBL, AB016077; BAA88823.1; ".
DR      InterPro: IPR005289; GTP-binding_dom.
DR      InterPro: IPR006073; GTP_OBG.
DR      InterPro: IPR002917; MMR_Hsrl.
DR      InterPro: IPR005225; Small-GTP.
DR      Pfam: PF01926; MMR_Hsrl. 1.
DR      PRINTS: PR00326; GTP_OBG.
DR      TIGRfams: TIGR00650; MG442; 2.
DR      TIGRfams: TIGR00231; small_GTP; 2.
SO      SEQUENCE      436 AA; 48601 MW; EC32CC8650DED090 CRC64;

```

Query Match	66.78;	Score 1476;	DB 2;	Length 436;
Best Local Similarity	62.88;	Pred. No. 3e-85;		
Matches 274;	Conservative 83;	Mismatches 77;	Indels 2;	Gaps 21

```

Oy 1 MTRPIVIAIGRPVWGKSTFENRIVGEERSIVEDPGYTRDIRIYSSGEMLHHDNIIDTCG 60
Db 1 MALPVAIYIGRPVWGKSALEPNRINGERISIVEDVGVTRRIRITTKAMMLNROSIIIDTCG 60
Oy 61 IETGDAPFOTOTBAQAEIATADEAVIIPWNVNREGLTQSDENVAQILYKSKRPVYLAVK 120
Db 61 IDVDVDAPEKQIKHQADIMTEADVIYVWSAKGICITDAOEYAKILYRTHKPEYIYLAVK 120
Oy 121 VDNKHEKRTDVIYDFESLGFCEPPIGSHGLCIGDLLVAWVSHGGEZEEDPYDEDTLRSLI 180
Db 121 VDNPEKRSATYDFETALGLDPIYPSASANGITGVDLAIYDNLPAEAOEE-SSDIITKPSL 179
Oy 181 IGRPNVKGSLVNAIIGEDRVIVSNVAGTTRDAIDTETS-YDGDVYLLIDTAGMRKRGKV 239
Db 180 IGRPNVKGSLNAIIGEDRVIASPVACTTRDAIDTETFEDEGEGLFIIDTAGMRKSGAV 239
Oy 240 YESTETKSYLRLALKALERSNVLYVIDEOGILIEDKRVAGYAHEDSKAVYIYVNRMDTV 299
Db 240 YETETKYSVYMRAPRAIDRSDIYLAWMLNEBIRYDKRIAGFPAHEAGKGLVYVNNWMDI 299
Oy 300 EKDSKTMKKFEDEVKREGEFLDYAOIAFVSKAKERTRLTLEPYINAESEHHKKRVQOSTL 359
Db 300 KKNRRIVAGMENDIDBNRQYIRYAIYEVSAVYQORLKLPRVYIKOVSQONRIRISSVL 359
Oy 360 NEVYVDAISNRPPTDKGRRLNVYATOVAAIEPPTVYFVNDVLEMLHFSYKRYLENOIRA 419
Db 360 NDVYVDAVAINPPTDKGRKRLKIFAYATOVSVKRPPTVYFVNEEELMHFSYLRLENDIRO 419
Oy 420 AFGEEGPIHIITAKKR 435
Db 420 AFVFEGETPIRLAKKR 435

```

RESULT 6	
Q8DS90	
ID Q8DS90	PRELIMINARY:
AC Q8DS90	PRT: 436 AA.
DT 01-MAR-2003 (TRENBLET, 23, Created)	
DT 01-MAR-2003 (TRENBLET, 23, Last sequence update)	

DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
 DE Phosphoglycerate dehydrogenase.  
 GN PGDA OR SMU.1920.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=UA159 / ATCC 700610 / Serotype C;  
 RA A.J. D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S.,  
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;  
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
 RT pathogen."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
 DR EMBL: AE015016; AAN59531.1;  
 KM Complete proteome.  
 SO SEQUENCE 436 AA; 48585 MW; 2408428A91C2A097 CRC64;

Query Match 66.5%; Score 1472; DB 16; Length 436;  
 Best Local Similarity 62.6%; Pred. No. 5.3e-85;  
 Matches 273; Conservative 84; Mismatches 77; Indels 2; Gaps 2;

OY 1 MTKPIVAIYGRPNVSKSTIFNRIGERSYIEDTPGVTRDRISSGENLTHDNIIDTGG 60  
 DB 1 MALPTVAIYGRPNVSKSLAFNRIGERSIYEDVEGVTTRITKAMLNROPSIIDTGG 60  
 OY 61 IETGAPFOTOIRAOAEIAIDAVIIFMNVNREGLTOSDEMAOILYKSKRPVLAIVK 120  
 DB 61 IODVAPFMEOKHODIAMTEADVIIVVSAKSGITDADEYAKILYRHKRVILAIVK 120  
 OY 121 VDNEMERTDYDFYSLGEGEPYISGSHGLGLDLDAVYSHGEEEDPYEDTIRLSI 180  
 DB 121 VDNEMERTDYDFYSLGEGEPYISGSHGLGLDLDAVYSHGEEEDPYEDTIRLSI 180  
 OY 181 ICRPNVGSLLVNAIIGEDRVIVSNVAGTTRADAIOTEXS-YGODVVLIDTGMKRRKQV 239  
 DB 180 ICRPNVGSLLVNAIIGEDRVIVSNVAGTTRADAIOTEXS-YGODVVLIDTGMKRRKQV 239  
 OY 240 YESTEKSYVALAKAIENSNNVLYVIADEGIIIDOKRVAGVIAHEGKAAVYVYVNMKV 299  
 DB 240 YESTEKSYVALAKAIENSNNVLYVIADEGIIIDOKRVAGVIAHEGKAAVYVYVNMKV 299  
 OY 300 EKDSKMKKFEDEVKKEFOFLDVAQIAVSAKERTRLTLFPIYINASENNKRVOSTL 359  
 DB 300 KDNRTVAQWEDAIKDNFQIPYAPIVFVSATVKORLKLDPVYKOVSOQNRIPSAVL 359  
 OY 360 NEVMTDAISMNPTPTDKGRNLNVYATOVAIERTPVFVNDVLELHNSFYKRYLENOIRA 419  
 DB 360 NOVVDVAIINPTPTDKGRNLNVYATOVAIERTPVFVNDVLELHNSFYKRYLENOIRA 419  
 OY 420 AFGEGCTPIHIAIKRR 435  
 DB 420 AFGEGCTPIHIAIKRR 435

## RESULT 7

08G58 PRELIMINARY: PRT: 445 AA.  
 AC 08G58:  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE GTP-binding protein (fragment).  
 OS Hellobacillus mobilis.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Hellobacteriaceae;  
 OC Hellobacillus.  
 NCBI\_TaxID=28064;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 DR MEDLINE=22337798; PubMed=12446909;

RA Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,  
 RA Blankenship R.E.;  
 RT "Whole-genome analysis of photosynthetic prokaryotes."  
 RL Science 298:1616-1620(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Liollos K.G., Chu L., Ostrovskaya O., Mendybaeva N., Koukharenko V.,  
 RA Gerdes S., Kyrides N., Overbeck R.;  
 RL Submitted (Aug-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AY142760; AAN87364.1;  
 FT NON-TER 445 445  
 SO SEQUENCE 445 AA; 49885 MW; 1D6A692E387AD61C CRC64;

Query Match 60.6%; Score 1342; DB 2; Length 445;  
 Best Local Similarity 56.8%; Pred. No. 8.8e-77;  
 Matches 250; Conservative 83; Mismatches 99; Indels 8; Gaps 3;

OY 1 MTKPIVAIYGRPNVSKSTIFNRIGERSYIEDTPGVTRDRISSGENLTHDNIIDTGG 60  
 DB 4 MAPPIVAVGRPNVSKSTIFNRIGERSYIEDTPGVTRDRISSGENLTHDNIIDTGG 63  
 OY 61 IETG--DAPFOGIRAOAEIAIDAVIIFMNVNREGLTOSDEMAOILYKSKRPVLAIV 118  
 DB 64 LEFGAGGNFSEVYIYKQAAIAEADVILFVADKAGITADDEVAALVLRKTPVFLV 123  
 OY 119 NKVDNEMERTDYDFYSLGEGEPYISGSHGLGLDLDAVYSHF--GEEEDPYEDT 175  
 DB 124 NKIEDSQDKYEFEPALGLDPIPSATKGMNGLDLDAVIELAPPEDEEDP---DT 180  
 OY 176 IRLSTIIGRPNVGSLLVNAIIGEDRVIVSNVAGTTRADAIOTEXSYGODVVLIDTGMK 235  
 DB 181 IRLSTIIGRPNVGSLLVNAIIGEDRVIVSNVAGTTRADAIOTEXSYGODVVLIDTGMK 240  
 OY 236 KGVYESTEKSYVALAKAIENSNNVLYVIADEGIIIDOKRVAGVIAHEGKAAVYVYVNM 295  
 DB 241 KGVYESTEKSYVALAKAIENSNNVLYVIADEGIIIDOKRVAGVIAHEGKAAVYVYVNM 300  
 OY 296 MDEVKDSKMKKFEDEVKKEFOFLDVAQIAVSAKERTRLTLFPIYINASENNKRVOSTL 355  
 DB 301 MDEVKDSKMKKFEDEVKKEFOFLDVAQIAVSAKERTRLTLFPIYINASENNKRVOSTL 360  
 OY 356 SSTLNEVMTDAISMNPTPTDKGRNLNVYATOVAIERTPVFVNDVLELHNSFYKRYLE 415  
 DB 361 TSVINELMTDILRVTPAPSDGRRLKILYVTSVKOPTFVFNDEBELHNSFYQHIN 420  
 OY 416 QIRAEFGCTPIHIAIKRR 435  
 DB 421 RRETFGCTPIHIAIKRR 440

## RESULT 8

08KH12 PRELIMINARY: PRT: 435 AA.  
 AC 08KH12:  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Conserved hypothetical GTP-binding protein.  
 OS Lactobacillus delbrueckii (subsp. bulgaricus).  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillus.  
 NCBI\_TaxID=1585;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 11842;  
 RA Serrif P., Deruy R., Ehrlich S.D., Maguin E.;  
 RT "Lactobacillus delbrueckii spp. bulgaricus hlda region."  
 RL Submitted (Apr-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AY094626; AAM22484.1;  
 DR InterPro: IPR000795; EF\_GTPbind.  
 DR InterPro: IPR005289; GTP-binding\_dom.  
 DR InterPro: IPR006073; GTP\_OBG.  
 DR InterPro: IPR002917; MMR\_HSR1.



QY 125 EMRT-DVYDFYSLGCEPPIPSGSHGLGDLDAVSHRGEEDDPIDETIRLSITGR 183  
DB 124 NKOSIDHEHLERFGRGKPFISANHGIGIGDLDLIVD---TKLKEKEEDTFRFCITGR 180  
QY 184 PNVKSSLVNALIGEDRVIVSVNACTRDAIDTEVSYDGODYVLIDTAGMKKKQVVEST 243  
DB 181 PNVKSSLVNACILNEDRMITSNIANTRDALIDSNEKKNLDLLIITDAGIRKCKIOENV 240  
QY 244 EKYSVLRALKAIERSNVVLVIDAEOGIIEDOKRQVAGYAHOGKAAVIVYNNKNDYV-EKD 302  
DB 241 DKXAYLRVQSGISRSRLIYIVLDGSEFNEQEDVYAGLAHKANIPSLIIVNNKNDIVKEKO 300  
QY 303 SKTKMKFFEDVRKEPFLDYAOIAFVSAKERTRLTLEPTYNESNNHKKRQVOSTLNEY 362  
DB 301 EKMNMKFTITIRIKFELWTPIVFLSALENKRIATLFEINETSIRENLNLKFNKILITDL 360  
QY 363 VTDALSNMPTPTDKGRNLNVFYATOVAIEPTFVYVNDVLMHFSYKRYLENOIRAFG 422  
DB 361 VFKLOMLNSPPLNRRGRKINHVTOVGOIPTFVLCNPEYLHFSYARLTENEIRKSLG 420  
QY 423 FECPPIHITARR 435  
DB 421 LNMVPTITLVFKNK 433

RESULT 11

Q8PF59 PRELIMINARY; PRT; 503 AA.

AC 08PF59  
DT 01-MAR-2003 (TReMBLrel. 23, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE Probable GTP-binding protein enga.  
GN C3033  
OS Escherichia coli O6.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
CX NCBI\_TaxID=217992;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-06:H1 / CPT073 / ATCC 700928;  
RX MEDLINE-22388234; PubMed-12471157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Rossch P.,  
RA Rasko D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,  
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
RA Mobley H.L.T., Domeneberg H.S., Blattner F.R.;  
RT "Extensive mosaic structure revealed by the complete genome sequence  
of uropathogenic Escherichia coli.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
DR EMBL; AE016764; AAN81483.1; -.  
KW Complete proteome.  
SQ SEQUENCE 503 AA; 56573 MW; 711A9BD7863280A CRC64;

Query Match 41.6%; Score 930; DB 16; Length 503;

Best Local Similarity .42.0%; Pred. No. 4.6e-50;

Matches 197; Conservative 91; Mismatches 143; Indels 38; Gaps 4;

QY 4 PIVAVGRPNVSKSTIFNRIYGERVSIEDTFCYTRDRIRYSSGEMLTGHDNIIDTGCIEI 63  
DB 16 PVVALVGRPNVSKSTIFNRLTRTRDALVADFGLTRDRKYGRAIEGREFICIDGIDG 75  
QY 64 GDAFPOTQIRAOAEIAIDEADYIFMVNVREGLOSDEMVAQIYKSKRPVLAIVNKNVDN 123  
DB 76 TEGCVETRNAQDSLAIEADYVLFMVNARAGLAPRADAIAKHLRSKKEFTFLVANKTDG 135  
QY 124 MEKRTDYDFYSLGCEPPIPSGSHGLGDLDAV----- 160  
DB 136 LDQDAVVDVFLAGEEYIPYIAASHGRGVLSLLEHVLTPMMDLAPQEEVDEDAEYWAQF 195  
QY 161 ---SHFGEEDDPYEDT--IRLSIIGRPVNGKSSLVNALIGEDRVIVSVNACTRDAID 215  
DB 196 EARENGEEDDPOPSLPIKALVGRPNVSKSTIFNRLIGERVVVYVNDPPTTDSIT 255

QY 216 TETSDODVYLLDTAGAKKKKGVYESTKYSVALAKAIERSNVVLVIDAEOGITEOD 275  
DB 256 IPMRIDREYVLLDTAGVRRKRTIDAVKEFVITLQAIEDANVVLVIDAREGCSIDDD 315  
QY 276 KRVAGVHEQKAVVIVYNNKNDVNEKSKTMKFFEDVRKEPFLDYAOIAFVSAKERTR 335  
DB 316 LSLGFLYNSGRSLIYVNNKMDLSQEVKEQKELDFR--LGTFIDFARVHIFSLHGSC 373  
QY 336 LRTFPIYNESNNHKKRQVOSTLNEVVDALSNMPTPTDKGRNLNVFYATOVAIEPTF 395  
DB 374 VGNLFESVRZAYOSSSTRPGTSMLTRITMAVEDHQPLVNGRKYKLTALAGGYNPYV 433  
QY 396 VFYVNDVLMHFSYKRYLENOIRAFGECPPIHIT-----IARRN 436  
DB 434 VTHGNQVKDLPDSYKRYLNVFRRSLDVGSPRIQFREGENPYANNKN 482

RESULT 12

Q8DF02 PRELIMINARY; PRT; 496 AA.

AC 08DF02  
DT 01-MAR-2003 (TReMBLrel. 23, Created)  
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE GTP binding protein.  
GN VY10423.  
OS Vibrio vulnificus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
CX NCBI\_TaxID=672;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-CMCP6;  
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
RA Choy H.E.;  
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";  
RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE016798; AAO08946.1; -.  
KW Complete proteome.  
SQ SEQUENCE 496 AA; 55469 MW; 2719F6BA312B7923 CRC64;

Query Match 40.6%; Score 898.5; DB 16; Length 496;

Best Local Similarity 42.4%; Pred. No. 1e-48;

Matches 195; Conservative 84; Mismatches 146; Indels 35; Gaps 4;

QY 4 PIVAVGRPNVSKSTIFNRIYGERVSIEDTFCYTRDRIRYSSGEMLTGHDNIIDTGCIEI 63  
DB 3 PVVALVGRPNVSKSTIFNRLTRTRDALVADFGLTRDRKYGQARVGEHDFIVDTGIDG 62  
QY 64 GDAFPOTQIRAOAEIAIDEADYIFMVNVREGLOSDEMVAQIYKSKRPVLAIVNKNVDN 123  
DB 63 SESEVERKMAQSLAIREADYVLFVNDGRALITSDAIAHLRIKIEKATMLVYNNKVDG 122  
QY 124 MEMTVDYDFYSLGCEPPIPSGSHGLGDLDAVSHF-----GEEBE--DPY 171  
DB 123 IDAASADFMQGVDEMVOIAAHGRGTALIERALDPFDDNLLSANNGEIEDLDMR 182  
QY 172 DEPT-----IRLSIIGRPVNGKSSLVNALIGEDRVIVSVNACTT 210  
DB 183 DEDAEOQYSEEDAEBSLRLDOPKILALITIGRRPNVSKSTLTNLTCEERVVYVNDPQGT 242  
QY 211 RDAIDTEYSVODGVYLLDTAGAKKKKGVYESTKYSVALAKAIERSNVVLVIDAEOG 270  
DB 243 RDSITITMERDGRVYLLDTAGVRRKRTIDAVKEFVITLQAIEDANVVLVIDAREG 302  
QY 271 IIEODKRVAGVHEQKAVVIVYNNKNDVNEKSKTMKFFEDVRKEPFLDYAOIAFVSA 330  
DB 303 ISDODLSLCLFALNAGRSLVLAIVNKNDCG--DNEVKEVWKELDRRLGPFDFARIFISA 360  
QY 331 KERRLRLTLPYVYNESNNHKKRQVOSTLNEVVDALSNMPTPTDKGRNLNVFYATOVAI 390  
DB 361 LHGTGVGHLEFESVQAEYRSATTRVGTSVLTRINKMALEDHQPVRVNRKRYKLTALAGGY 420



[illegible]



Qy 62 -----EIGDAPFOTQIRAOEIAIDADVITFMVNRG 95  
 Db 217 KSPGCVMEELNSTTIGMCGITLSREAAIARMSHMIKQTAANDSAVITFVVDGAG 276  
 Qy 96 LTOSDEMAOIL--YKSKRPVLAIVKNVDN-MEMRTDVFYSLGFGEPYISGSHGL 152  
 Db 277 PSQADVEIADWLKRYSHYIILAVNKCSPKGLMQASEFWSGFP-TPIDISLSTGCT 335  
 Qy 153 GDLDAVVS-----HFGSEEDPYDEDTIRLSIGRPNGKSSLVNATIGEDRVY 203  
 Db 336 GELDLVSGGLKLEIMENIEESEEEN--IPAAIIGRPNGKSSILNATVREDRTY 392  
 Qy 204 SNAGCTRAIDTEVS-YGODYVLIDTAGMKKCKYVS--TEKISVLKALKALERN 259  
 Db 393 SPVSGTTRADIDELTFGPCKEKFRLIDTAGIKKSSVASGSGSTEASVNRARAIIRD 452  
 Qy 260 VLVVVIDAEGGIIEDDKRVAGYAHGOKAVVIVVKNKMDY--EKDSTMKKFEDEVKREFO 318  
 Db 453 VVALVIEAMACITEODLKIAERIEREGKGLVAVVKNKMDYIPNKHQETAAYHEDVREKLR 512  
 Qy 319 FLDYAQIAFVSAKERTRLTLPYINEASENHKKVOSTLTENVYTAISM-NPTPTDNG 377  
 Db 513 SLNMAPIYSTAIGSVNIVYAAATVOKERSRLSTAILNOVIRCAVAFKSPPTRG 572  
 Qy 378 RLNVFATQVAIEPPTFVFNVDVLAHFSYKRYLENOIRAFGFCPTPIHIAKR 435  
 Db 573 KRCRVYCTOAIIRPTFVFVFNDAKLFSDTYRRYMEROLRTDAGFAGTPIRLMRSR 630

## RESULT 15

OREC36  
 ID OREC36 PRELIMINARY: PRT: 487 AA.

AC 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE GTP-binding protein Enga.  
 GN ENG4 OR SO3308.  
 OS Shewanella oneidensis.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 CC Alteromonadaceae; Shewanella.  
 CX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MR-1;  
 RX MEDLINE-22297686; Pubmed-12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 Read T.D., Eisen J.A., Seshadri R., Ward N., Meche B., Clayton R.A.,  
 Meyer T., Tsaplin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 Vamathevan J., Weidman J., Imprial M., Lee K., Berry K., Lee C.,  
 Mueller J., Khouri H., Gill J., Ueberback T.R., McDonald L.A.,  
 Reddylynn T.V., Smith H.O., Venter J.C., Nelson K.H., Fraser C.M.,  
 "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 Shewanella oneidensis."  
 RT Natl. Biotechnol. 20:1118-1123(2002).  
 RL EMBL: AEO15769; AAN56306.1;  
 DR TIGR: SO3308;  
 KW Complete proteome.  
 SO SEQUENCE 487 AA; 54531 MW; B75F8C6CA1BF9C20 CRC64;

Query Match 38.8%; Score 859; DB 16; Length 487;  
 Best Local Similarity 39.98; Pred. No. 3; le-46;  
 Matches 180; Conservative 98; Mismatches 147; Indels 26; Gaps 4;

Qy 4 PIVAVGRPNVSKSTIFNRIVGERSIVEDTFCVTRDRIRYSSGEMTHDNFIIDFGIEI 63  
 Db 3 PVALVGRPNVSGSTLFRNLRTDALVADPGLTRDRKXGKRAFLSGYEFTIVDGTGIDG 62  
 Qy 64 GDAPFOTQIRAOEIAIDADVITFMVNRGSLTOSDEMAVQIYKSKRPVLAIVKNVDN 123

Db 63 TEESIEFKMAEQSLAAIEADVYVFWTDRAGLTADLSINQHLRSNQKTFVYANKIDG 122  
 Qy 124 MEMRTDVFYSLGFGEPYISGSHGLGDLDAVVSF-----GEEF---EDPY 171  
 Db 123 IDADSACERFWSLGLLEVYQMAAOGKGTNMIETALPVAEANGIERGEEVEDEKRY 182  
 Qy 172 DEE-----TIRLSIGRPNGKSSLVNATIGEDRVYSNVAGCTRAIDTEVS 219  
 Db 183 TEDEAEAEOKRLQDLPIKILAIIGKPNVGSSTLTNILLCEERVVYDDEPGTTRDSIYIME 242  
 Qy 220 YGODYVLIDTAGMKKCKYVSLEKYSVLKALKALERSNVVLAVIDAEGGIIEDDKRVA 279  
 Db 243 RDGREYIYIDTAGVRRSKVHEYIEKFSYIKLAVDEANVYLLIDARCGVAEODLGL 302  
 Qy 280 GYAHGOKAVVIVVKNKMDYVERKDSKTKMKFEDEVKREFOVDYQAIAFVSAKERTRLRTL 339  
 Db 303 GFALNAGRALYIVVKNKMDGIDGIR--DRVNSELDORRIGFTDFARHIFISALHGTGCHL 360  
 Qy 340 FPYINEASENHKKVOSTLTENVYTAISMNPTPTDKGRNLNVFATQVAIEPPTFVFN 399  
 Db 361 FESIEEAYDSATRRVSTSLMTRIMQMSQDDHQPLVNGRRAVKLYKAHAGGYNPPIYVHG 420  
 Qy 400 NDVELMHFSYKRYLENOIRAFGFCPTPIH 430  
 Db 421 NOVSKLPDSYKRYMNYFRRLKLVYGTPIQL 451

Search completed: September 10, 2003, 00:35:04  
 Job time : 105 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 17:16:39 : Search time 129.915 Seconds  
(without alignment)  
8041.284 Million cell updates/sec

Title: US-09-815-242-1463  
Perfect score: 387  
Sequence: 1 gatctctctctctctccca.....ttcaccgtgaataataac 387

Scoring table: IDENTITY\_NUC  
Gap 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	387	100.0	387	23	AA548886
2	387	100.0	1311	22	AA548886
3	387	100.0	1311	23	AA548886
4	387.8	99.2	1305	23	AA548886
5	383.8	99.2	1311	23	AA548886
6	327	84.5	3621	18	AA548886
7	295	76.2	372	23	AA548886
8	262.2	67.8	1332	24	AA548886

9	262.2	67.8	3269	22	AA548886	S. epidermidis gen
10	234	60.5	298	23	AA548886	Staphylococcus aur
11	234	60.5	298	23	AA548886	Staphylococcus aur
12	205.8	53.2	319630	24	AB067194	Listeria innocua C
13	205.8	53.2	2944528	24	AB067194	Listeria monocytog
14	205.8	53.2	3011208	24	AB067194	Listeria innocua D
15	174.4	45.1	1311	24	AB067194	Bacillus lichenif
16	162.6	42.0	1308	24	AB067194	Streptococcus poly
17	162.6	42.0	2355589	24	AB067194	Streptococcus poly
18	160.4	41.4	1308	24	AB067194	Streptococcus poly
19	160.4	41.4	2155561	24	AB067194	Streptococcus poly
20	153	39.5	1311	21	AA545516	Staphylococcus aur
21	151.4	39.1	1308	25	AB074747	S. pneumoniae type
22	151.4	39.1	1311	21	AA545516	Streptococcus pneu
23	151.4	39.1	5066	19	AA545516	Streptococcus pneu
24	151.4	39.1	2162598	25	AB067194	Streptococcus pneu
25	146	37.7	245	23	AA548886	Staphylococcus aur
26	101	26.1	960	22	AA548886	S. epidermidis ope
27	89.4	22.1	1512	23	AA548886	E. coli DNA for ce
28	89.4	22.1	11574	22	AA548886	DNA encoding novel
29	86	22.2	157	23	AA548886	Staphylococcus aur
30	84.4	21.8	1512	22	AA548886	Haemophilus influe
31	84.4	21.8	1512	22	AB064943	DNA encoding haemo
32	84.4	21.8	1515	23	AA548886	Haemophilus influe
33	83.8	21.7	1473	23	AA548886	Salmonella typhi D
34	80.6	20.8	640681	24	AB067194	Buchnera sp. genom
35	80.2	20.7	1830121	17	AA548886	Haemophilus influe
36	73.2	18.9	1370	19	AA548886	S. pneumoniae derl
37	73.2	18.9	33140	22	AA548886	Genomic fragment
38	69.8	18.0	1146	25	AA548886	C. glutamicum derl
39	69.8	18.0	1557	22	AA548886	C. glutamicum codin
40	69.8	18.0	349980	22	AA548886	C. glutamicum codin
41	69.8	18.0	349980	22	AA548886	C. glutamicum codin
42	64	16.5	25360	22	AA548886	S. spinoza DNA fire
43	64	16.5	29736	22	AA548886	S. spinoza DNA fire
44	62	16.0	580073	18	AA548886	Mycoplasma genital
45	61.4	15.9	78845	21	AA548886	N. meningitidis pa

## ALIGNMENTS

RESULT 1	AA548886	standard; DNA; 387 BP.
XX	AA548886	
AC	AA548886	
XX	AA548886	
DT	13-FEB-2002	(first entry)
XX	13-FEB-2002	
DE	Staphylococcus aureus cellular proliferation inhibitory sequence #110.	
XX	Staphylococcus aureus	
XX	antisense; ss: prokaryotic cellular proliferation;	
KW	antibiotic; antibacterial; drug design.	
XX	Staphylococcus aureus.	
OS	Staphylococcus aureus.	
PN	WO200170955-A2.	
XX	27-SEP-2001.	
PD	27-SEP-2001.	
XX	21-MAR-2001; 2001WO-US09180.	
PF	21-MAR-2001; 2000US-191078P.	
XX	21-MAR-2001; 2000US-191078P.	
PR	23-MAY-2000; 2000US-206848P.	
PR	26-MAY-2000; 2000US-207727P.	
PR	23-OCT-2000; 2000US-242578P.	
PR	27-NOV-2000; 2000US-253625P.	
PR	22-DEC-2000; 2000US-257931P.	
PR	16-FEB-2001; 2001US-263308P.	
XX	(ELIT-) ELITRA PHARM INC.	
PA	(ELIT-) ELITRA PHARM INC.	
XX	(ELIT-) ELITRA PHARM INC.	



QY 301 TTGGATGTCATCACCATTTCATACACCTGTATCATATATTTGAATCATGTG 360  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 208 TTGGATGTCATCACCATTTCATACACCTGTATCATATATTTGAATCATGTG 149  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 361 TTACCATTCACCTGAGAAATTAATAC 387  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 148 TTACCATTCACCTGAGAAATTAATAC 122

## RESULT 3

AAS54997/c  
 ID AAS54997 standard; DNA: 1311 BP.

AAS54997;  
 13-FEB-2002 (first entry)

Staphylococcus aureus DNA for cellular proliferation protein #1309.

Antisense: ds; prokaryotic cellular proliferation gene;  
 antibiotic; antibacterial; drug design.

Staphylococcus aureus.

WO200170955-A2.

27-SEP-2001.

21-MAR-2001: 2001WO-US09180.

21-MAR-2000: 2000US-191078P.

23-MAY-2000: 2000US-206848P.

26-MAY-2000: 2000US-207727P.

23-OCT-2000: 2000US-242578P.

27-NOV-2000: 2000US-253625P.

22-DEC-2000: 2000US-257931P.

16-FEB-2001: 2001US-269308P.

(ELIT-) ELITRA PHARM INC.

Haselbeck R, Ohlgen KL, Zyckind JW, Wall D, Trawick JD, Carr GJ;  
 Yamamoto RT, Xu HH;  
 WPI: 2001-611495/70.  
 P-PSDB: AAU37138.

New polynucleotides for the identification and development of  
 antibiotics, comprise sequences of antisense nucleic acids -  
 Claim 27; Seq ID No 8634; 511pp; English.

The invention relates to antisense inhibitors of genes essential to  
 prokaryotic cellular proliferation, their use in identifying the  
 genes, their use in the discovery of novel antibiotics, the essential  
 genes themselves and the encoded proteins. The prokaryotes used are  
 Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 invention is also useful for the identification of potential new targets  
 for antibiotic development. The antisense nucleic acids can also be used  
 to identify proteins used in proliferation, to express these proteins,  
 and to obtain antibodies capable of binding to the expressed proteins.  
 The proteins can be used to screen compounds in rational drug discovery  
 programmes. The antisense nucleic acid sequence is also useful to screen  
 a wide variety of organisms. The present sequence encodes an  
 essential prokaryotic cellular proliferation protein.  
 Note: The sequence data for this patent did not form part  
 of the printed specification, but was obtained in electronic  
 format directly from WIPO at  
 ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1311 BP: 452 A: 184 C: 278 G: 397 T: 0 other:

Query Match 100.0%; Score 387; DB 23; Length 1311;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-98;  
 Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCTCTTCCTCTTACCAAAATGAGAAACATGATATACAGTCACCAAGACCTA 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 508 GATCTCTTCCTCTTACCAAAATGAGAAACATGATATACAGTCACCAAGACCTA 449  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 61 AACCATGACCCCTGATATCGGATACGGTTCCACCAATCCTAATGATAGAAATCA 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 448 AACCATGACCCCTGATATCGGATACGGTTCCACCAATCCTAATGATAGAAATCA 389  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 121 CGTCTGACCATTTTCATATTAATTAATTAATTAATTAATTAATTAATTAAT 180  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 388 CGTCTGACCATTTTCATATTAATTAATTAATTAATTAATTAATTAATTAAT 329  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 181 ATTTGATATAATTTGACGACCATTTTCATATTAATTAATTAATTAATTAAT 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 328 ATTTGATATAATTTGACGACCATTTTCATATTAATTAATTAATTAATTAAT 269  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 241 CCATATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 268 CCATATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 209  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 301 TTGGATGTCATCACCATTTCATACACCTGTATCATATATTTGAATCATGTG 360  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 208 TTGGATGTCATCACCATTTCATACACCTGTATCATATATTTGAATCATGTG 149  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 361 TTACCATTCACCTGAGAAATTAATAC 387  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 148 TTACCATTCACCTGAGAAATTAATAC 122

## RESULT 4

AAS51646/c  
 ID AAS51646 standard; DNA: 1305 BP.

AAS51646;  
 13-FEB-2002 (first entry)

Staphylococcus aureus DNA for cellular proliferation protein #63.

Antisense: ds; prokaryotic cellular proliferation gene;  
 antibiotic; antibacterial; drug design.

Staphylococcus aureus.

WO200170955-A2.

27-SEP-2001.

21-MAR-2001: 2001WO-US09180.

21-MAR-2000: 2000US-191078P.

23-MAY-2000: 2000US-206848P.

26-MAY-2000: 2000US-207727P.

23-OCT-2000: 2000US-242578P.

27-NOV-2000: 2000US-253625P.

22-DEC-2000: 2000US-257931P.

16-FEB-2001: 2001US-269308P.

(ELIT-) ELITRA PHARM INC.

Haselbeck R, Ohlgen KL, Zyckind JW, Wall D, Trawick JD, Carr GJ;  
 Yamamoto RT, Xu HH;  
 WPI: 2001-611495/70.  
 P-PSDB: AAU37187.

New polynucleotides for the identification and development of  
 antibiotics, comprise sequences of antisense nucleic acids -

PS Claim 27; Seq ID No 4228; 511pp; English

CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Mycobacteria*  
CC *tuberculosis*, *Pseudomonas aeruginosa* and *Mycobacterium leprae*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

**50** Sequence 1305 BP; 447 A; 181 C; 277 G; 400 T; 0 other;

Query Match	99.28;	Score 383.8;	DB 23;	Length 1305;
-------------	--------	--------------	--------	--------------

Matches 385; Conservative 0; Mismatches 2; Indels 0; Gaps 0.

[illegible]XX  
PD 27-SEP-2001

PF 21-MAR-2001; 2001WO-US09180.

PR	21-MAR-2000	2000US-191078P
PR	22-MAY-2000	2000US-206848P
PR	26-MAY-2000	2000US-207732P
PR	23-OCT-2000	2000US-242578P
PR	27-NOV-2000	2000US-253625P
PR	22-DEC-2000	2000US-257931P
PR	16-FEB-2001	2001US-269308P

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlson KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX

DR P-PSDB; AAU37006.

PT New polynucleotide

XX

PS Claim 27; Seq ID No 8502; 511pp; English

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence encodes an essential prokaryotic cellular proliferation protein.

**SD** Sequence 1311 BP; 451 A; 181 C; 277 G; 402 T; 0 other.

Query Match	99.2%	Score 383.8	DB 23	Length 1311
-------------	-------	-------------	-------	-------------

Matches	385;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0
---------	------	--------------	----	------------	----	--------	----	------	---

Oy	1	GATCTCTTCTCTCTCAACAAATGGAACACATGCATCTAACCAAGTCACCAACACTA	60
Db	508	GATCTTCTTCTCTTCAACAAATGGAACACATGCATCTAACCAAGTCACCAACACTA	4459
Oy	61	AACCATGTGACCCTGATCGATGCGGTACCGTTCACCAATCTAATGAAATGAATATACA	1205
Db	448	AACCATGTGACCCTGATCGATGCGGTTCACCAATCTAATGAAATGAATATACA	3899
Oy	121	GCTGTGACGATTTTCATATATCTACTCTTTGTAACCGTAATACAGCGGTTTTAG	1800
Db	388	GCTGTGACGATTTTCATATATCTACTCTTTGTAACCGTAATACAGCGGTTTTAG	3299
Oy	181	ATTGTATATAAATTGAGGACCATTTTCATCGCTTTGTGTCAATCTCTTACGACGTTAA	2410
Db	328	ATTGTATATAAATTGAGGACCATTTTCATCGCTTTGTGTCAATCTCTTACGACGTTAA	2699
Oy	241	CCATATAAATAATATACATCGGCTTCATATGCGGATTTCTGCTCGCGCTTAATTTGTG	3000
Db	268	CCATATAAATAATATACATCGGCTTCATATATGAGGGAATTTCTGCTCGCGCTTAATTTGTG	2099
Oy	301	TTTGGATGTGTCATCACCAAATTTGCAATACACCTGTATCAATATATATGAANCAATGTG	3600

```

Db      208 TTTGGATGTGATCCCAATTCAATCCACTGATCAATTAATGAAATCATGTG 149
Qy      361 TTAACATTCACCGAGATTAATAC 387
Db      148 TTAACCACTCACCGAAGATTAATAC 122

RESULT 6
AAV74669/C
ID      AAV74669 standard; DNA; 3621 BP.
XX
XX      AAV74669;
XX
XX      16-MAR-1999 (first entry)
XX
XX      Staphylococcus aureus contig SEQ ID #358.
XX
XX      Computer readable medium; vaccine; S.aureus infection; immunodetection;
XX      cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
XX      skin infection; surgical wound infection; scalded skin syndrome;
XX      toxic shock syndrome; ds.
XX
XX      Staphylococcus aureus.
XX
XX      Key      Location/Qualifiers
XX      misc_feature 481..540
XX      /tag=8
XX      /note="these bases represent a line of missing text in
XX      the sequence listing in the specification. They
XX      are included to maintain the nucleotide numbering
XX      given in the specification for this DNA sequence"
XX
XX      misc_feature 2281..2340
XX      /tag=b
XX      /note="these bases represent a line of missing text in
XX      the sequence listing in the specification. They
XX      are included to maintain the nucleotide numbering
XX      given in the specification for this DNA sequence"
XX
XX      EP786519-A2.
XX
XX      30-JUL-1997.
XX
XX      07-JAN-1997; 97EP-0100117.
XX
XX      05-JAN-1996; 96US-0009861.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX
XX      Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
XX      Rosen CA;
XX
XX      WPI; 1997-374922/35.
XX
XX      Polynucleotide(s) and proteins derived from Staphylococcus aureus
XX      stored on computer readable medium and used in the production of
XX      anti-S.aureus vaccines
XX
XX      Claim 1: Page 1241-1243; 3271pp; English.
XX
XX      This sequence represents one of 5191 Staphylococcus aureus DNA sequences
XX      of the invention. The DNA sequences are recorded on a computer readable
XX      medium, preferably selected from a floppy or hard disk, random access
XX      memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
XX      the S.aureus DNA sequences allows putative functions to be assigned so
XX      that protein-encoding or regulatory regions of commercial, therapeutic or
XX      industrial importance can be obtained. Specifically, sequences which are
XX      likely to encode antigens have been identified and these polypeptides can
XX      be used in a vaccine composition against S.aureus infection. The
XX      polypeptides can also be used in a kit for the immunodetection of
XX      S.aureus in a sample. S.aureus is implicated in numerous human diseases,
XX      including cellulitis, eyelid infections, food poisoning, osteomyelitis,
XX      skin and surgical wound infections, scalded skin syndrome, toxic shock

```

```

CC      syndrome, etc. Organisms transformed with the DNA sequences can be used
CC      for recombinant production of the polypeptides. The new DNA sequences
CC      (and their fragments) are useful as primers or probes for isolating
CC      homologues of any of the S.aureus DNA sequences contained on the
CC      computer readable medium.
XX
XX      Sequence 3621 BP; 1279 A; 442 C; 692 G; 1085 T; 123 other:
XX
XX      Query Match      84.5%; Score 327; DB 18; Length 3621;
XX      Best Local Similarity 84.5%; Pred. No. 1.6e-81;
XX      Matches 327; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
XX
Qy      1 GATCTTCCTCCCTTCACCAAAATGAGAAACAATGCATCTAACAAGTCCACCAAGCTTA 60
Db      2637 GATCTTCCTCCCTTCACCAAAATGAGAAACAATGCATCTAACAAGTCCACCAAGCTTA 2578
Qy      61 AACCATGTGACCCGATATGCGATACGCTTCACCAAAATGCTAATGAAATGATACA 120
Db      2577 AACCATGTGACCCGATATGCGATACGCTTCACCAAAATGCTAATGAAATGATACA 2518
Qy      121 CGTCTGACGATTTCCATATATCTACTTTGTAACCGCTAATAGACCGGTTTTTAG 180
Db      2517 CGTCTGACGATTTCCATATATCTACTTTGTAACCGCTAATAGACCGGTTTTTAG 2458
Qy      181 ATTTGTAATTAATTTGACGACGACATTCATCGCTTGTGTGCAATCCTTCAGCAGCTTA 240
Db      2457 ATTTGTAATTAATTTGACGACGACATTCATCGCTTGTGTGCAATCCTTCAGCAGCTTA 2398
Qy      241 CCATATAAATAATACATCCGCTTCATCTATGCGCATTTCTGCTGCGCTTAATTTGTG 300
Db      2397 CCATATAAATAATACATCCGCTTCATCTATGCGCATTTCTGCTGCGCTTAATTTNN 2338
Qy      301 TTTGGAATGTGTCATCAACCAATTCATACACCTGTATCAATATTAATGAAATCATGTG 360
Db      2337 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2278
Qy      361 TTAACCATTCACCGAAGATTAATAC 387
Db      2277 TTAACCATTCACCGAAGATTAATAC 2251

RESULT 7
AAS50706
ID      AAS50706 standard; DNA; 372 BP.
XX
XX      AAS50706;
XX
XX      13-FEB-2002 (first entry)
XX
XX      Staphylococcus aureus cellular proliferation inhibitory sequence #1930.
XX
XX      antisense; ss; prokaryotic cellular proliferation;
XX      antibiotic; antibacterial; drug design.
XX
XX      Staphylococcus aureus.
XX
XX      WO200170955-A2.
XX
XX      27-SEP-2001.
XX
XX      21-MAR-2001; 2001WO-US09180.
XX
XX      21-MAR-2000; 2000US-191078P.
XX      23-MAY-2000; 2000US-206848P.
XX      26-MAY-2000; 2000US-207272P.
XX      23-OCT-2000; 2000US-242578P.
XX      27-NOV-2000; 2000US-253625P.
XX      22-DEC-2000; 2000US-257931P.
XX      16-FEB-2001; 2001US-269308P.
XX
XX      (ELITR-) ELITRA PHARM INC.
XX
XX      Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Travick JD, Carr GJ;

```

PI Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

PT New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids -

PS Claim 1: Seq ID No 3283; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence is an antisense  
CC oligonucleotide of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 372 BP; 113 A; 83 C; 51 G; 125 T; 0 other;

Query Match Best Local Similarity 76.2%; Score 295; DB 23; Length 372;

Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 93 CCAATCTTAATGATAGAAATCATACAGCTGTGACGATTCATATATCTACTTGG 152  
1 CCAATCTTAATGATAGAAATCATACAGCTGTGACGATTCATATATCTACTTGG 60  
OY 153 TTAACCGCTAATAGACGCGCTTTTGTAGATTGATATTAATTTGAGCGACATTCATCG 212  
DB 61 TTAACCGCTAATAGACGCGCTTTTGTAGATTGATATTAATTTGAGCGACATTCATCG 120  
OY 213 CTTTGTGCAATCTTACGACGCTTAACCAATAAATAAATTAACATCCGCTTCATCTATG 272  
DB 121 CTTTGTGCAATCTTACGACGCTTAACCAATAAATAAATTAACATCCGCTTCATCTATG 180  
OY 273 GCGATTCTGCGCTGCTTAATTTGTTGGAATGTCATCACCATTCAATATCAATACCA 332  
DB 181 GCGATTCTGCGCTGCTTAATTTGTTGGAATGTCATCACCATTCAATATCAATACCA 240  
OY 333 CCGTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 387  
DB 241 CCGTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 295

RESULT 8

ID ABR90883/c standard: DNA: 1332 BP.

XX ABR90883;

XX 24-JUL-2002 (first entry)

XX *Staphylococcus epidermidis* ORF nucleic acid sequence SEQ ID NO:346.

XX *Staphylococcus epidermidis*; open reading frame: ORF; bacterial infection;

XX antibacterial; gene therapy; gene; ds.

XX *Staphylococcus epidermidis*.

XX US6380370-B1.

PD 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

XX 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI: 2002-381255/41.

XX P-PSDB: ABR38338.

XX Disclosure: SEQ ID 346; 267pp; English.

CC ABR90538 to ABR93374 represent *Staphylococcus epidermidis* open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABR95124 to ABR97960. The *S. epidermidis* sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly *S. epidermidis* infections. The sequences can be used to  
CC screen for compounds able to interfere with the *S. epidermidis* life  
CC cycle or inhibit *S. epidermidis* infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.

XX Sequence 1332 BP; 465 A; 190 C; 267 G; 410 T; 0 other;

Query Match Best Local Similarity 67.8%; Score 262.2; DB 24; Length 1332;

Matches 309; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

OY 1 GATCTTCCTCCCTTCACCAAAATGAGAAACGATGATCAACAGTCACCAAGACCTA 60  
DB 529 GATCTTCCTCCCTTCATTTATTAAGTTTCACCAACTGCATCTAGCAATTCACAGTCTA 470  
OY 61 AACCATGTGACCTGATATCGGATAGAGGTTACCAAAATCTAATGATAGAAATCATAC 120  
DB 469 ATCCATGTGACCAAGAAATAGATGATGATCTCCAAAGCCATTAAGAAATCATAC 410  
OY 121 CGTCTGACGATTTTCATATATCTACTTGTATTAACCGTAATAGACCGGTTTGTAG 180  
DB 409 TATCATTTAGCAATTTCAAGATATCACTTATTCACAGCAATTAACAGGTTTCTTAG 350  
OY 181 ATTGTATTAATTAATGAGGAGCAATTCATGCGCTTTGTGCAATCTTCACGACGCTTA 240  
DB 349 ATTGTATTAATTAATGAGGAGCAATTCATGCGCTTTGTGCAATCTTCACGACGCTTA 290  
OY 241 CCAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTGTC 300  
DB 289 CCAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTGTC 230  
OY 301 TTTGGAATGTCATCACCATTCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 360  
DB 229 TTTGGAATGTCATCACCATTCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 170  
OY 361 TTAACCATTCACCTGAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 387  
DB 169 TTAACCATTCACCTGAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 143

RESULT 9

ID AAH54708 standard: DNA: 3269 BP.

XX AAH54708;

XX 03-SEP-2001 (first entry)





Query Match 60.5%: Score 234: DB 23: Length 298:  
 Best Local Similarity 100.0%: Pred. No. 7.9e-56:  
 Matches 234: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 GATCTTCTTCCTTCACCAAAATGAGAAACATGTCATTAACAATGACCAAGACCTA 60  
 DB 65 GATCTTCTTCCTTCACCAAAATGAGAAACATGTCATTAACAATGACCAAGACCTA 124  
 OY 61 AACCATGTGACCGTATTCGATACGCTTCACCAAAATGAGAAATGATGATCA 120  
 DB 125 AACCATGTGACCGTATTCGATACGCTTCACCAAAATGAGAAATGATGATCA 184  
 OY 121 CGTCTGTACGATTCATATATCTACTTTGTTAACCGGTAATAGACCGGTTTTAG 180  
 DB 185 CGTCTGTACGATTCATATATCTACTTTGTTAACCGGTAATAGACCGGTTTTAG 244  
 OY 181 ATTGTATATAAATTGAGCGACCATTCATCGCTTTGTGTCAATCCCTTCACGCA 234  
 DB 245 ATTGTATATAAATTGAGCGACCATTCATCGCTTTGTGTCAATCCCTTCACGCA 298

# RESULT 11

AAS50723 ID AAS50723 standard: DNA: 298 BP.

AC AAS50723:

DT 13-FEB-2002 (first entry)

DE Staphylococcus aureus cellular proliferation inhibitory sequence #1947.

KM Antisense; ss: prokaryotic cellular proliferation;

KM antibiotic; antibacterial; drug design.

OS Staphylococcus aureus.

PN W0200170935-A2.

PD 27-SEP-2001.

PR 21-MAR-2001: 2001MO-US09180.

PR 21-MAR-2001: 2000US-191078P.

PR 26-MAY-2000: 2000US-206848P.

PR 23-OCT-2000: 2000US-207727P.

PR 27-NOV-2000: 2000US-243578P.

PR 22-DEC-2000: 2000US-253625P.

PR 16-FEB-2001: 2001US-269308P.

PA (ELIT-) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr CJ;

PI Yamamoto RT, Xu HH;

DR WPI: 2001-611495/70.

PS Claim 1: Seq ID No 3300: 511pp: English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen CC for homologous nucleic acids which are required for cell proliferation in CC a wide variety of organisms. The present sequence is an antisense CC oligonucleotide of the invention.  
 CC Note: The sequence data for this patent did not form part CC of the printed specification, but was obtained in electronic CC format directly from Wipo at CC ftp.wipo.int/pub/published\_Pct\_sequences.

SO Sequence 298 BP: 91 A: 69 C: 43 G: 95 T: 0 other:

Query Match 60.5%: Score 234: DB 23: Length 298:  
 Best Local Similarity 100.0%: Pred. No. 7.9e-56:  
 Matches 234: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 GATCTTCTTCCTTCACCAAAATGAGAAACATGTCATTAACAATGACCAAGACCTA 60  
 DB 65 GATCTTCTTCCTTCACCAAAATGAGAAACATGTCATTAACAATGACCAAGACCTA 124  
 OY 61 AACCATGTGACCGTATTCGATACGCTTCACCAAAATGAGAAATGATGATCA 120  
 DB 125 AACCATGTGACCGTATTCGATACGCTTCACCAAAATGAGAAATGATGATCA 184  
 OY 121 CGTCTGTACGATTCATATATCTACTTTGTTAACCGGTAATAGACCGGTTTTAG 180  
 DB 185 CGTCTGTACGATTCATATATCTACTTTGTTAACCGGTAATAGACCGGTTTTAG 244  
 OY 181 ATTGTATATAAATTGAGCGACCATTCATCGCTTTGTGTCAATCCCTTCACGCA 234  
 DB 245 ATTGTATATAAATTGAGCGACCATTCATCGCTTTGTGTCAATCCCTTCACGCA 298

# RESULT 12

AB067194/c ID AB067194 standard: DNA: 319630 BP.

AC AB067194:

DT 29-AUG-2002 (first entry)

DE Listeria innocua contig DNA sequence #7.

KM Antibacterial; Listeria; food contamination; mutational analysis;

KM infection; ds.

OS Listeria innocua.

PN W0200228891-A2.

PD 11-APR-2002.

PR 04-OCT-2001: 2001MO-FR03061.

PR 04-OCT-2000: 2000FR-0012697.

PA (INSP ) INST PASTEUR.

PA (CNRS ) CNRS CENT NAT RECH SCI.

PI Kunst F, Glaser P;

DR WPI: 2002-332479/37.

New genomic sequences from *Listeria* species, useful for detection, treatment and prevention of infection, also related polypeptides, antibodies and modulators -  
 Claim 5: SEQ ID 7: 180pp: French.

The present invention relates to nucleic acid sequences (AB067194-AB071213) from *Listeria* sp. The sequences are useful as probes and primers for identification and/or detection of *Listeria* (e.g. as contaminants in foods, or mutational analysis) and for analysis of gene expression. Proteins encoded by the nucleic acid sequences can be

CC used to screen for compounds that modulate gene expression, replication  
 CC and pathogenicity of *Listeria* (potential therapeutic agents), also for  
 CC creating infections by *Listeria*, and are useful as immunogens in  
 CC anti-*Listeria* vaccines.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SO Sequence 319630 BP; 105207 A; 55428 C; 66726 G; 92263 T; 6 other:

Query Match 53.2% Score 205.8; DB 24; Length 319630;  
 Best Local Similarity 70.9% Pred. No. 5.5e-47;  
 Matches 273; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

OY 3 TCTTCTCCTCTTCCACCAATATGAGAAACATGCTATACAGTACAGCAAGCTTAA 62  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 51123 TCGCTTCTCTTTTGGAAATATGACGACAGCATCAAGCATACACAGCCCTAGT 51064  
 OY 63 CCATGTGACCTGATATCGGATACGGTTACCAATCTTAATGATAGAAATCATACAG 122  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 51063 CCATGAGAACCGAAATTTGATAGCGCTCACCAACCGAAAGAAATAAAGTCATATAATC 51004  
 OY 123 TCTGTACCATTTCCATATTAATCTACTTGTCTTAACCGCTAATACGACCGGTTTTTAAAT 182  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 51003 TGATCAGCATTTCTGCGTATCTACTTATTAATCGCTAAACATTTGTTTATTAGAC 50944  
 OY 183 TTGTATAAATTTGAGGAGGACCATTTCAATCGCTTGTGTAATCGCTTACAGCAAGCTTAAAC 242  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 50943 CGGTAAAGAAATTTTGTCTACTGTCTGTGCTATCGCTACGTAACCCCTTACAGCAAGCTTAAAC 50884  
 OY 243 ATAAATATATACATCCGCTTCACTATGCGCAATTTGCTGCGCTCTAATTTGTGT 302  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 50883 ATAAATATATATACATCCGCTTCACTATGCGCAATTTGCTGCGCTCTAATTTGTGT 50824  
 OY 303 TGAATGCTGATCAGCAATTTCAATACGCTGTATCAATATATATGAATCATGTGT 362  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 50823 AAGATGCTGCTGCGAAAGATCAATACGCTGTATCAATATATATGAATCATGTGT 50764  
 OY 363 AACCATTCACCTGAAAGATTAATAC 387  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 50763 AGCCATTCGCTGAAATTAATATGC 50739

## RESULT 13

AB03041  
 ID ABA03041 standard; DNA: 2944528 BP.

AC ABA03041:

XX 05-FEB-2002 (first entry)

XX *Listeria* monocytogenes EGD-e genome sequence.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KM vitamin B12; bacterial infection; disease; ds.

XX *Listeria* monocytogenes.

XX WO200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

XX (INSP) INST PASTEUR.

PI Buchrieser C, Frangoul L, Couve E, Rusniok C, Eslih H, Dehoux P;  
 PI Dusjurgel O, Checouant F, Nedjari H, Glaser P, Kunst F, Cossart P;  
 PI Danileis J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 PI Chakraborty T, Domann E, Hain T, Berche P, Chablit A, Durant L;

PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 PI Medueno E, De Pablos B, Wehlund J, Kaerst U, Entlan K, Hauf J;  
 PI Rose M, Voss H;  
 XX  
 DR WPI; 2002-010914/01.

XX Genomic sequence for *Listeria* monocytogenes, useful e.g. for treatment  
 PT and prevention of *Listeria* and related bacterial infections, and  
 PT related polypeptides

XX Claim 1: SEQ ID No 1: 192pp: French.

XX The present sequence is the genome sequence of *Listeria* monocytogenes  
 CC EGD-e. This sequence and fragments of this sequence are useful for  
 CC selecting probes and primers for detecting genes in *L. monocytogenes* and  
 CC related organisms, and to study genetic polymorphisms and other genomes.  
 CC proteins (AB847297-AB850149) expressed from the present sequence are  
 CC useful for raising specific antibodies, identification of *L.*  
 CC monocytogenes and related organisms, and for biosynthesis and  
 CC biodegradation, especially biosynthesis of vitamin B12. This sequence and  
 CC proteins encoded by it are also useful for selecting compounds that  
 CC regulate gene expression and cell replication and modulate *L.*  
 CC monocytogenes-related diseases. In addition, this sequence and proteins  
 CC encoded by it are useful in pharmaceutical and vaccine compositions for  
 CC the treatment or prevention of infections by *L. monocytogenes* and related  
 CC organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

SO Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other:

Query Match 53.2% Score 205.8; DB 24; Length 2944528;  
 Best Local Similarity 70.9% Pred. No. 1.1e-46;  
 Matches 273; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

OY 3 TCTTCTCCTCTTCCACCAATATGAGAAACATGCTATACAGTACAGCAAGCTTAA 62  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 2011933 TCGCTTCTCTTTTGGAAATATGACGCGACAGCATCAAGCATATGCTATAAAT 2011992  
 OY 63 CCATGTGACCTGATATCGGATACGGTTACCAATCTTAATGATAGAAATCATACAG 122  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 2011993 CCATGTGACCGAAATTTGATAGCGCTCACCAACCGAAAGAAATAAAGTCATATAAT 2012052  
 OY 123 TCTGTACCATTTCCATATTAATCTACTTGTCTTAACCGCTAATACGACCGGTTTTTAAAT 182  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 2012053 TGATCAGCATTTCTGCGTATCTACTTATTAATCGCTAAACATTTGTTTATTAGAC 2012112  
 OY 183 TTGTATAAATTTGAGCGACCATTTGCTTGTCTCAATCTTACAGCAAGCTTAAAC 242  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 2012113 CGGTAAAGAAATTTTGTCTACTGTCTGTGCTACGTAACCTCTTACAGCAAGCTTAAAC 2012172  
 OY 243 ATAAATATATACATCCGCTTCACTATGCGCAATTTGCTGCGCTCTAATTTGTGT 302  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 2012173 ATAAATATATATACATCCGCTTCACTATGCGCAATTTGCTGCGCTCTAATTTGTGT 2012232  
 OY 303 TGAATGCTGATCAGCAATTTCAATACGCTGTATCAATATATATGAATCATGTGT 362  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 2012233 AAAAATGCTCATTCGGAAGATCAATACACCTGTATCAATATATGAATCATGTGT 2012292  
 OY 363 AACCATTCACCTGAAAGATTAATAC 387  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 DB 2012293 AGCCATTCGCTGAAATTAATATGC 2012317

## RESULT 14

AB069245  
 ID ABO69245 standard; DNA: 3011208 BP.

AC ABO69245:

XX 29-AUG-2002 (first entry)



Sat, Sep 13 12:47:59 2003

us-09-815-242-1463.rng

Page 11

Db	489	AAAAATTGGCGCTGACCGCGCTCAAGACGAGATCCGCGAGNCCAAAGGCGCATGCGTCCATGAAAT	430
Qy	80	CGGATACGGTTACCCAAATCTTAATGAAATAGAAATCATACGCTCTGTACGCAATTCAT	139
Db	429	CGGGAAGGGCTGCGCCCAAAGCCGAGGCGATAAAAAATCCGTAATGTTTCTCTCATTTTCAGG	370
Qy	140	ATTATCTACGTTCTGTTAAACCGTAAATACGACGCGCTTTTAAATTTGTAATAAATTTGAGC	199
Db	369	ATTATTCACATTATTAACCGCCGACGACGACGAGGTTTTTTTGTACGCTATTAATATTTTCGC	310
Qy	200	GACCAATTCATGCGTTTGTCAATCTTCAACGCAAGTAACATATAAATAATTAATACATC	259
Db	309	CACCTTCTTATCAAGCGCGTGTACGCCCTTCCGCGCGCTTGCATGAAAAATTAATACATC	250
Qy	260	CGCTTCATCATATGCGCATTTCTGCGTGGCTCTAATTTGTGTGGAAATGATGATCACAC	319
Db	249	TGCTTTCTTCATGCGCATCTCGGCTGTGATGGCGAATCTGCGCCAAAACGGCTCATGCGC	190
Qy	320	AAATTCAAATACCACTGTATCAATAAATTTGAAATCATGTTAACCAATTCACCTGAGA	379
Db	189	GACCTTGAAATCCCGCCGTATCAATACAGGTTGGAAGCGTGAATTCGATTCAGCACCTGGAGAGCT	130
Qy	380	ATAAATAC 387	
Db	129	GTAATATCC 122	

Search completed: September 12, 2003, 17:32:09  
Job time : 138.915 secs

Sat Sep 13 12:47:59 2003

us-09-815-242-1463.rge

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 17:17:09 : Search time 1580.02 Seconds

(without alignments)  
10020.134 Million cell updates/sec

Title: US-09-815-242-1463

Perfect score: 387

Sequence: 1 gatctctctctctcacca.....ttcacctgaagaataaac 387

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

GenBank: 1: gb\_ba: 2: gb\_ba: 3: gb\_ba: 4: gb\_ba: 5: gb\_ba: 6: gb\_ba: 7: gb\_ba: 8: gb\_ba: 9: gb\_ba: 10: gb\_ba: 11: gb\_ba: 12: gb\_ba: 13: gb\_ba: 14: gb\_ba: 15: gb\_ba: 16: gb\_ba: 17: gb\_ba: 18: gb\_ba: 19: gb\_ba: 20: gb\_ba: 21: gb\_ba: 22: gb\_ba: 23: gb\_ba: 24: gb\_ba: 25: gb\_ba: 26: gb\_ba: 27: gb\_ba: 28: gb\_ba: 29: gb\_ba: 30: gb\_ba: 31: gb\_ba: 32: gb\_ba: 33: gb\_ba: 34: gb\_ba: 35: gb\_ba: 36: gb\_ba: 37: gb\_ba: 38: gb\_ba: 39: gb\_ba: 40: gb\_ba: 41: gb\_ba:

Prod. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	387	100.0	1308	AX622668	AX622668 Sequence
2	387	100.0	301550	AP003134	AP003134 Staphylococcus aureus
3	387	100.0	333750	AP004827	AP004827 Staphylococcus aureus
4	387	100.0	346900	AP003362	AP003362 Staphylococcus aureus
5	262.2	67.8	3269	AF270032	AF270032 Staphylococcus aureus
6	262.2	67.8	3269	AX145350	AX145350 Sequence
7	262.2	67.8	300892	AE016747	AE016747 Staphylococcus aureus
8	210.6	54.4	290117	AE017028	AE017028 Staphylococcus aureus
9	209	54.0	304680	AE017002	AE017002 Staphylococcus aureus
10	205.8	53.2	313450	AL596170	AL596170 Staphylococcus aureus
11	205.8	53.2	319630	AX413016	AX413016 Sequence
12	205.8	53.2	347050	AL591981	AL591981 Staphylococcus aureus
13	205.8	53.2	349980	AX417046	AX417046 Sequence
14	205.8	53.2	349980	AX641670	AX641670 Sequence
15	198.4	51.3	24887	BACSERA	BACSERA Sequence
16	198.4	51.3	213680	BSUB0012	BSUB0012 Staphylococcus aureus
17	191.4	49.5	302173	AE016951	AE016951 Staphylococcus aureus
18	188.2	48.6	300550	AP004599	AP004599 Staphylococcus aureus
19	185	47.8	300550	AP001512	AP001512 Staphylococcus aureus
20	182.4	47.1	302050	AL935257	AL935257 Staphylococcus aureus
21	174.4	45.1	1311	AX433884	AX433884 Sequence
22	162.6	42.0	11071	AE006309	AE006309 Staphylococcus aureus
23	162.6	42.0	12434	AE006498	AE006498 Staphylococcus aureus
24	162	41.9	1311	AX607165	AX607165 Sequence
25	162	41.9	44145	AX602195	AX602195 Sequence
26	160.4	41.4	174050	SAG766852	SAG766852 Staphylococcus aureus
27	159.4	41.2	52276	AE014265	AE014265 Staphylococcus aureus
28	159.4	41.2	323825	AP005146	AP005146 Staphylococcus aureus
29	157.8	40.8	12370	AE009978	AE009978 Staphylococcus aureus
30	157.8	40.8	12370	AE008523	AE008523 Staphylococcus aureus
31	153	39.5	12340	AX570314	AX570314 Sequence
32	151.4	39.1	1308	AE008523	AE008523 Staphylococcus aureus
33	151.4	39.1	5066	AR218847	AR218847 Sequence
34	151.4	39.1	5066	BD003759	BD003759 Polynucleotide
35	151.4	39.1	10310	AE007464	AE007464 Staphylococcus aureus
36	151.4	39.1	151947	SPNEU1902	SPNEU1902 Staphylococcus aureus
37	151.4	39.1	349980	AX571764	AX571764 Sequence
38	151.4	39.1	349980	AX571765	AX571765 Sequence
39	148.8	38.7	13860	AE015016	AE015016 Staphylococcus aureus
40	148.2	38.3	3737	AB016077	AB016077 Staphylococcus aureus
41	139.4	36.0	301278	AE015939	AE015939 Staphylococcus aureus
42	125.2	32.4	296750	AP003191	AP003191 Staphylococcus aureus
43	117	30.2	10861	AE007680	AE007680 Staphylococcus aureus
44	109.2	28.2	3557	AY094626	AY094626 Lactobacillus
45	101	26.1	960	AX144037	AX144037 Sequence

# ALIGNMENTS

RESULT 1  
AX622668/c AX622668 1308 bp DNA linear PAT 20-FEB-2003  
DEFINITION Sequence 5631 from Patent WO02094868.  
ACCESSION AX622668  
VERSION AX622668.1 GI:28450653  
KEYWORDS  
SOURCE  
ORGANISM Staphylococcus aureus  
Staphylococcus aureus  
Bacteria; Firmicutes; Bacillales; Staphylococcus.  
REFERENCE  
1 Masignani,V.C., Mora,M.C. and Scarselli,M.C.  
TITLE Staphylococcus aureus proteins and nucleic acids  
JOURNAL Patent: WO 02094868-A 5631 28-NOV-2002;  
Chilton Spa (IT)

```

FEATURES
    source
        Location/Qualifiers
            1..1308
                /organism="Staphylococcus aureus"
                /mol_type="genomic DNA"
                /db_xref="taxon:1280"
BASE COUNT      450 a      184 c      278 g      396 t
ORIGIN
Query Match      100.0% Score 387; DB 6; Length 1308;
Best Local Similarity 100.0%; Pident. No. 2e-77; 0; Indels 0; Gaps 0;
Matches 387; Conservative 0; Mismatches 0;

QY 1 GATCTCTCTCTCTCTACCAAAATGAGAACAACATGCATTAACAAGTCACCAAGCCTA 60
Db 508 GATCTCTCTCTCTCTACCAAAATGAGAACAACATGCATTAACAAGTCACCAAGCCTA 449
QY 61 AACCATGTGACCCGTGATATCGGATACGGTTCCACCAATCTTATGATGAAATCATACA 120
Db 448 AACCATGTGACCCGTGATATCGGATACGGTTCCACCAATCTTATGATGAAATCATACA 389
QY 121 CGTCTGTACGCATTTCCATATTACTTCTTAAACCGCTAATACGACCGGTTTTTAG 180
Db 388 CGTCTGTACGCATTTCCATATTACTTCTTAAACCGCTAATACGACCGGTTTTTAG 329
QY 181 ATTTGTATAAATTTGACGACGACATTTTCATCGCTTTGTGTCATCCTGCACGACGTTAA 240
Db 328 ATTTGTATAAATTTGACGACGACATTTTCATCGCTTTGTGTCATCCTGCACGACGTTAA 269
QY 241 CCATATAAATAATATACATCGCTTCATATGCGGATTTGCGCTGCGCTTAATTTGTG 300
Db 266 CCATATAAATAATATACATCGCTTCATATGCGGATTTGCGCTGCGCTTAATTTGTG 209
QY 301 TTGGATATGATGATACCATTTCAATACACCTGATACATTAATTTGAATCATGTG 360
Db 208 TTGGATATGATGATACCATTTCAATACACCTGATACATTAATTTGAATCATGTG 149
QY 361 TTAACCATTCACCTGAGAGATAATATAC 387
Db 148 TTAACCATTCACCTGAGAGATAATATATAC 122

RESULT 2
LOCUS      AP003134      301550 bp      DNA      linear      BCT 24-APR-2003
DEFINITION Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
ACCESSION  AP003134 BA000018
VERSION     AP003134.2 GI:14349226
KEYWORDS
SOURCE
ORGANISM   Staphylococcus aureus subsp. aureus N315
REFERENCE  Bacteria: Firmicutes; Bacillales; Staphylococcus.
AUTHORS    Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,
            Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Imai, J., Ito, T., Kanamori, M.,
            Matsunuma, H., Maruyama, A., Murakami, H., Hosoyama, A.,
            Mizutani, O., Y., Takahashi, N. K., Sawano, T., Inoue, R., Kato, C.,
            Sekimizu, K., Hiraoka, H., Kuwano, S., Goto, S., Yabuzaki, J.,
            Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C.,
            Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.
            Whole genome sequencing of methicillin-resistant Staphylococcus
            aureus
            Lancet 357 (9264), 1225-1240 (2001)
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (30-JAN-2001) Director-General, Biotechnology Center,
National Institute of Technology, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/,

```

```

COMMENT      Tel:81-3-3481-1933, Fax:81-3-3481-8424)
FEATURES
    source
        On Jun 12, 2001 this sequence version replaced gi:13701256.
        Location/Qualifiers
            1..301550
                /organism="Staphylococcus aureus subsp. aureus N315"
                /mol_type="genomic DNA"
                /strain="N315"
                /sub_species="aureus N315"
                /db_xref="taxon:158879"
                /complement(1249..947)
                /gene="SA1293"
                /note="ORFID:SA1293"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="BAB42553.1"
                /db_xref="GI:13701259"
                /translation="MSIFPMITFYLPALPMAOHKYSNTERYSQVRSCTGR
                EVALLEHANGIYDEVVKGFLDHDYDRKKVSLSPANDRPSVAGTAIAHEVG
                HAIDHQGYWELFRFAALVPANLSSLSYIMILYITLIGSAGSFTALDIAGLM
                SLAVRFLVTLPEFEDASSRAMKQITALINVEKEYKHARKVLSAAATMYAATAVAV
                ASLVRIILARSDD"
                /complement(1001..1588)
                /gene="SA1294"
                /complement(1001..1588)
                /note="ORFID:SA1294"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="BAB42554.1"
                /db_xref="GI:13701260"
                /translation="MTINFQWYLYLQYRSLMMLILCNIGMIGYIYWGDSHTPM
                HFRIYDPSPALIYLVISILSILOKNSIIEALFVLFKGIWAVINLVFIHQ
                GDITNGVLVMSHSIMAVOALFYPRFRSMIGISVAMIMWFLNDYIDYFILQPPY
                DRTTHVMQIGVLSCLSVFGLLIYELKRLCK"
                /complement(1578..2153)
                /gene="SA1295"
                /complement(1578..2153)
                /note="ORFID:SA1295"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="BAB42555.1"
                /db_xref="GI:13701261"
                /translation="MSFTLNQIKESFTEYLLFQYRKRSLAVWLVYIKVNEAKIAN
                HVPDRKINHNETLETAIVGSNASHIOFTKRNITLNTNPFIYIANHNCAEDIOHFA
                NYSKREORLDLYVAOLNIESPSTOTLHDSMAIDRRHALLDILHNRDLSLQNH
                EKORFVLTOLNLTFLVKNHNFEDLADD"
                /complement(2167..3411)
                /gene="SA1296"
                /complement(2167..3411)
                /note="ORFID:SA1296"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="BAB42556.1"
                /db_xref="GI:13701262"
                /translation="MEDYKILIDINQIKENLIDSNVNEAITTDNDALFLIGETLYN
                FQLMPQGLEVRVLYHKYPOBELIYFIEGLMSENGTDEALEYLSVSPERKMLEE
                ADLYOQINNMESVAIDKLQELALEPNDPIIHFALEMLYDQGLRAVSEYVLELGG
                EYOVNGVNLFSRMADCSLQSGNSDAIRLVLEINDEMTSEDEYLKRAISYKNDITGE
                AIKIMTLLSDPDYIOGYLYLQSLYENENYPDALETGKBGLRSOFYKELMTWTCG
                LEIHGDANBGLVLLKQALEVDNAYGEPCLISLDYLRNEEDYEAILELTYVDEEDLD
                PTFMHLAFAYGGEERKQEAHPPELAYPTMKTNIDPMSDYFYFLIEIGQEQDAITLL
                NQLELPSNENMHDELSSLOS"
                /complement(3418..4716)
                /gene="atca"
                /complement(3418..4716)
CDS
gene
CDS

```



```

CDS                                     /gene="ndk"  
                                           complement(7777..8226)  
                                           /gene="ndk"  
                                           /note="ORFID:SA1301"  
                                           /codon_start=1  
                                           /transl_table=11  
                                           /product="nucleoside diphosphate kinase"  
                                           /protein_id="BAB82562.1"  
                                           /db_xref="GI:13701268"  
                                           /translation="MERTFLMIKRPDAVONRLIGEVISRIERKGLVLGYKALQVPKML  
                                           AETHNGEHOQKPFPNDLSIFTSAPFAVAWVGEDAVANSHIISTSPSASGCSIR  
                                           GOLGLTVRNIITSDLESAREINILMFENBIDITSASPDMHLYE"  
                                           complement(8318..9277)  
                                           /gene="gerCC"  
                                           complement(8318..9277)  
                                           /gene="gerCC"  
                                           /note="ORFID:SA1302"  
                                           /codon_start=1  
                                           /transl_table=11  
                                           /product="heptaprenyl diphosphate synthase component II"  
                                           /protein_id="BAB82563.1"  
                                           /db_xref="GI:13701269"  
                                           /translation="MAKLNNNEIKRYVEDLEKAIRSKDYSYLEOASLHLSSGGKRVPR  
                                           PAFVLLSOPFKDEOTSEDTYGVAAVALHELIHMATLVHDVIDKSQRKRCKTIIKKMD
```

Query Match 100.0% Score 387; DB 1; Length 301550;  
Best Local Similarity 100.0%; Pred. No. 1,2e-77;  
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1	GATCTTCCTCTCTCTCACCAAAATAGAAAACACTGCATCTAACAAGTCACCAGACCTTA	60
Db	13271	GATCTTCCTCTCTCTCACCAAAATAGAAAACACTGCATCTAACAAGTCACCAGACCTTA	13333
Oy	61	AACCATGTACCCCTATATCGATAGCGTTACACCAAATCCTATGAATGAATCATCA	120
Db	13331	AACCATGTACCCCTATATCGATAGCGTTACACCAAATCCTATGAATGAATCATCA	13339
Oy	121	CCTCTGTACGCATTTCATATTAATCTACTTTTGTAACCGCTAATACGACCGTTTTTGA	180
Db	13391	CCTCTGTACGCATTTCATATTAATCTACTTTTGTAACCGCTAATACGACCGTTTTTGA	13455
Oy	181	ATTTGTATATAATTTGAGCGACATTTTCATCGCTTTGTGTCAATCCTTCACGACGTTAA	240
Db	13451	ATTTGTATATAATTTGAGCGACATTTTCATCGCTTTGTGTCAATCCTTCACGACGTTAA	13511
Oy	241	CCATAAAATATAATACATCGCTTCATCTATAGGCAATTCGTGCTGCGCTTAATTTGTG	300
Db	13511	CCATAAAATATAATACATCGCTTCATCTATAGGCAATTCGTGCTGCGCTTAATTTGTG	13571
Oy	301	TTTGGATGTGCATACCAATTTTCATATCCACCTGTATCATATATATTTGAAATCATGTG	360
Db	13571	TTTGGATGTGCATACCAATTTTCATATCCACCTGTATCATATATATTTGAAATCATGTG	13631
Oy	361	TTAACCATTCACCTGAAGAATTAATAC	387
Db	13631	TTAACCATTCACCTGAAGAATTAATAC	13657

RESULT\_3  
AP004827  
LOCUS  
DEFINITION  
Staphylococcus aureus subsp. aureus MN2, complete genome,  
strain:WM2, section 6/10.  
AP004827 BA0000033  
ACCESSION  
VERSION  
KEYWORDS  
AP004827.1 GI:21204509  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
1 Baba,T., Takeuchi,F., Kuroda,M., Yuzawa,H., Aoki,K., Oguchi,A.,  
Nagai,Y., Iwama,N., Asano,K., Naito,T., Kuroda,H., Cui,L.,  
Yamamoto,K. and Hiramatsu,K.

TITLE Genome and virulence determinants of high virulence  
JOURNAL Community-acquired MRSA  
MEDLINE Lancet 359 (9320), 1819-1827 (2002)  
PUBMED 12044378  
2 (bases 1 to 333750)  
REFERENCE Director-General, Biotechnology Center, Aoki, K., Oguchi, A.,  
AUTHORS Nagai, Y., Asano, K., Iwama, N., Baba, T., Kuroda, M., Hiramatsu, K. and  
Kikuchi, H.  
DIRECT SUBMISSION  
TITLE Submitted (06-MAR-2003) Director-General, Biotechnology Center,  
JOURNAL National Institute of Technology and Evaluation, Biotechnology  
Center, 2Chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan  
(E-mail: bio@nite.go.jp, URL: http://www.bio.nite.go.jp/  
Tel:81-3-3481-1933, Fax:81-3-3481-8424)  
FEATURES  
source Location/Qualifiers  
1.333750  
/organism="Staphylococcus aureus subsp. aureus MW2"  
/mol\_type="genomic DNA"  
/strain="MW2"  
/db\_xref="taxon:196620"  
complement(278. .619)  
/gene="MW1341"  
complement(278. .619)  
/gene="MW1341"  
/note="ORFID: MW1341"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAB95206.1"  
/db\_xref="GI:21204510"  
/translation="MTNITEVFDNDEQEDIVPAKKKTLGSEARPIYDEYHAIID  
EKYQNDIGNIDFDKLMQIFVPMNKERIDMIEKTHFGIROMKTFKELINMA  
SYQRLNKKQV"  
complement(624. .1283)  
/gene="nth"  
complement(624. .1283)  
/gene="nth"  
/note="ORFID: MW1342"  
/codon\_start=1  
/transl\_table=1  
/product="endonuclease-like protein"  
/protein\_id="BAB95207.1"  
/db\_xref="GI:21204511"  
/translation="MWSKKALEMIDIVANMPDAECLEKHNPELTIAVLLSAOCT  
DVLNRYVTELEFKYKPEDYLVASDEELMNDIRSIGLRNKKRKLCSLIDQFN  
GEIPQTRKELESAGVGRKATVAVSVADERSLAVDTHVERVSRGICINRKNQVRO  
VEDRLCSVIPRDMNNSHOLIFGCRHCLARRPKDICPLLEDCREGQKRYKASLKE  
A"  
complement(1273. .1959)  
/gene="MW1343"  
complement(1273. .1959)  
/gene="MW1343"  
/note="ORFID: MW1343"  
hypothetical protein, similar to chromosome replication  
initiation protein dnab"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB95208.1"  
/db\_xref="GI:21204512"  
/translation="MDKQYKARPVIRRELDHYSDGLDEODLVLLKLIVASETS  
NKQSTELQKSGMRODITVONLORLELLEOKDEGRTPEYMNIDPEPEKL  
SHLKQOSMETKQNSKEKFKOLFRIELETTPARPSPREIETLNMIDVDHDSAIIO  
AALDEAHSLLKLSFKTMDRITLLMKNNKNTIDSDSKIREKFKMKHTVKTVPKD  
WLNGLNDGK"  
complement(2287. .3579)  
/gene="asnS"  
complement(2287. .3579)  
/gene="asnS"  
/note="ORFID: MW1344"  
/codon\_start=1  
/transl\_table=1  
/product="asparaginyl-tRNA synthetase"  
/protein\_id="BAB95209.1"  
/db\_xref="GI:21204513"  
/translation="MKTITKQAKDHLNDVYTGAVLTJNNRSSGKIAPILQLRDGTGFMQ  
GVVVSVEDEEVFLAKEITQESSLYVTGTTEDNSDGLGYEMQVKSIEVSEADHP  
ITPKNHGTETLDRHRLMSKKQHAVKIRNEVIAKATEYEPFNKQGFYKVPPLTAS  
APEGSELFHTKRYFDQDAFLSQSGQLLEAANAQKVSFGPTPRASKSTRHLLLE  
FWMIGEMAFTHNAESLEIQOYVTHVKSYLEKSLKLEIKERTSKLEKYPAPFPR  
ISYDAIEFLKSGEFDIEMGEDGAPHEGIAHHNDLVPFTYNNPTKIKFQYQPNP  
EHEETVLCADLIAPEGYEITGSGSERDQLELLEDPKREHGEDEBAVYIYDLNRYS  
VPHGFGGLERTYANISGVPHVETAPFRLNRLKP"  
complement(3901. .6594)  
/gene="ding"  
complement(3901. .6594)  
/gene="ding"  
/note="ORFID: MW1345"  
/codon\_start=1  
/transl\_table=1  
/product="probable ATP-dependent DNA helicase ding"  
/protein\_id="BAB95210.1"  
/db\_xref="GI:21204514"  
/translation="MGMAITVAVDLETGCGNLDPDITDITGTPYRNQGITDYHNSI  
RTNLEIPPEIQLTSTENMLQAPYFNQVADITDKINDCTFVANHVDLNTTKA  
FKDCNIQYPRKVIDTLEIFRIAPDPKSYQSELEAGITLVAHRAADDAATYAK  
LMILAPEKFEKPLDTLKLQYLSKQDKYLDVDFEVNRODARPLDKSYKEEOII  
YRKQYDFKPTTNGSLSKSLYSKAVDQGLTYRPOQLAETIIDLQMLSEKAMIEA  
SLGSGKSLAYLLAALNIEGKHWMISTRTYRLOSLENDIIPAMNALLFKXIAL  
IKSKSDYISGLISQILKDDTSNVEVILKMOQLIWTETPEGDIOELNLKGGQWVF  
DQKIEYVPAHVDVYNNPIKRNMONIIGITMHAHLHSDVENSIIYOLFDCIYDVA  
HRLPYALNOYTNELSYADIKYOLGLGKNNKMLKALIDOLEKRIELEKIDAPIID  
FGKASMEIHELMEQLSPRTITINDSOYDODIHRPNPPTPKOILDIAIID  
KLKRTLEFNKISIKTYKSLRKQILYLKQKFRKIDSLKAGTSISTKINSOKSTR  
LIVKDIAYKDLVIRQYERKSLFISGLTRKHSPEAKQLEPNQVFNFTVNTSL  
QSAKTSVFIFSDVASTQYKNDIVASIVSIIETTYITSSCLVLETSFQWANGIKCVH  
DMLNELPEDEDVVLTOONONKRIQOQFEDKALILDTSTFSGSDFQANGIKCVH  
IAKLFPNKNHAKYMLADSESTSEFEVYLPAAVIRFQGLRLRLRSFENDGIIYSFD  
DRLINSNYKNFEQTELENYRQKKGDIDQFGKLLRQYOKKK"  
complement(6618. .7589)  
/gene="MW1346"  
complement(6618. .7589)  
/gene="MW1346"  
/note="ORFID: MW1346"  
/codon\_start=1  
/transl\_table=1  
/product="truncated hypothetical protein"  
/protein\_id="BAB95211.1"  
/db\_xref="GI:21204515"  
/translation="MSKYSDVYQLLYKKNPNYISGOSIAESLNSRTAVKAYIDQK  
LEGKIDSVNHHGHLQLOLPDIYOCIIQYKSSALFDSFVYDSIDSTOLAAKSL  
VGNOSFEILSDEQTKGRNFRNHSMSGGGLMMSVNLPRVAFMSIKFNLTALG  
IRDAIOHFSODEEVKVPNDIYIDNCKVCGFLTEVANNNDGTEAIIIGIGINTQOE  
NDSISRRRATSVOLKDKNLDRYOFLEBLIEIKRYNPLTLPFSIRREYIAASN  
IWNRTILFTENDNQFKCAJDLDYDGLIVRDEAGSHRLISAQIDP"  
complement(7576. .8778)  
/gene="MW1347"  
complement(7576. .8778)  
/gene="MW1347"  
/note="ORFID: MW1347"  
hypothetical protein, similar to poly(A) polymerase"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB95212.1"  
/db\_xref="GI:21204516"  
/translation="MDSLSLEQARPLBEDIONGFEATVYGGSVADYMGKNTIIDT  
TTSATPDEISFSHTIPVGRKCHTINNVFNEND EYTFRREEDYDHRPSCVTV  
RDLYEQRODFTMAIAMOTAYKRLDYDGGQDINNIIIRVCGIAERREVEDLRLNI  
RCLEFDSQSLSDIATETFEAMRIOMADIFLSIEIVELTRKLNKINVEKSPHMLS  
LKAFNMPYFELDMNOINVTBAIDELIAIVSKFINDINSFLKLSNRQVNDIIO  
YIOINNALPSITTEQOLKHFVYDTHLIRKNVAAVAKANDIOGHPPLVNLQITD  
ETLRLRPLHNKDKMNVGVLMAHLNAKSGPLKXVDYLRQIEIAIVTGVNSEETILK  
WVDNHVKI"  
complement(8783. .9925)  
/gene="MW1348"  
complement(8783. .9925)

```

/gene="MW1348"
/note="ORFID:MW1348
hypothetical protein, similar to lipopolysaccharide
biosynthesis-related pr homolog"
/codon_start=1
/transl_table=11
/protein_id="BAB95213.1"
/db_xref="GI:21204517"
/translacion="MKIGTCTPSCGSGIITELGCIKIAENGHEVHTSNIPRI
KPLEMIFHOVENQYAFQJPPYDITLSTKIAEYIKETGDLIMHYAVHAIGIL
AREMSKIDIMWTLLHGTDTVAGDHSQGAIKGEIKSDIVTSVSKIAQETHEII
ETNKEIIPYINPARENFPKHTALKSQFGIAPEKVIHVSINROVKRIDIIEIF
AKYREKIPSKLILLDGPELPMKQLKLELNEEDVLFKQDCVSEFYQASDVLL
SEKESFGLTLEAMKTVGIGSNAGIKKEIKHGSGFVVDGCDSDASDAIRLLE
DKVLYNKLQKNMLADIAERGSSELTIDQVEYIYOKKLNHNKSKE"
/complement(10167..10484)
/gene="MW1348"
/complement(10167..10484)
/gene="MW1349"
/note="ORFID:MW1349"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB95214.1"
/db_xref="GI:21204518"
/translacion="MKSVMQREVDYIGQFQYFSPILANLRTPEVGLAREIN
HTYERKKKSEANITKAELGDNLFVLLCLANSNGIDMTESFNETMEKFNTRDKNR
ERK"
/complement(10820..11518)
/gene="MW1350"
/complement(10820..11518)
/gene="MW1350"
/note="ORFID:MW1350"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB95215.1"
/db_xref="GI:21204519"
/translacion="MSITFTIYFYILNVLPMQHKRSYERYSOVRSQGTGR
EVALEILHANGITDVEYKGFDTQHPKRVSLSPANDRPSVAGTAAAEVGS
HAIDHOGYFLRRALPVAANLSSLSLMIIMGLIITAIAGSAGSTALMIGGLM"

Query Match      100.0%: Score 387; DB 1; Length 333750;
Best Local Similarity 100.0%: Pred. 1.2e-77;
Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GATCTCTCTCTCTCAACAAATGAGAAACAATGCACTAATACCAAGTCAACCAAGCTA 60
DB      23844 GATCTCTCTCTCTCAACAAATGAGAAACAATGCACTAATACCAAGTCAACCAAGCTA 23903
QY      61 AACATGTGACCTGATATCGATACGGTTCACCAATCCTAATGAAATGAAATCATACA 120
DB      23904 AACATGTGACCTGATATCGATACGGTTCACCAATCCTAATGAAATGAAATCATACA 23963
QY      121 CGTGTGATGACGATTCATATTAATCTACTTGTGTAACCGGTATATACGACGGTGTGTTAG 180
DB      23964 CGTGTGATGACGATTCATATTAATCTACTTGTGTAACCGGTATATACGACGGTGTGTTAG 24023
QY      181 ATTGTATTAATTTGACGACATTTGATCGCTTGTGTCAATCCTTACGACGCTTAA 240
DB      24024 ATTGTATTAATTTGACGACATTTGATCGCTTGTGTCAATCCTTACGACGCTTAA 24083
QY      241 CCATTAATAATTAACATCGCTTCATCTATGCGGATTTCTGCGCTGCGCTTAATTTGTG 300
DB      24084 CCATTAATAATTAACATCGCTTCATCTATGCGGATTTCTGCGCTGCGCTTAATTTGTG 24143
QY      301 TTGTGAATGATGACCAATTTGATACCACTGATATCAATTAATTAATTAATTAATG 360
DB      24144 TTGTGAATGATGACCAATTTGATACCACTGATATCAATTAATTAATTAATTAATG 24203
QY      361 TTAAACATTCACTGAGAAATTAATAC 387
DB      24204 TTAAACATTCACTGAGAAATTAATAC 24230

```

```

RESULT 4
AP003362
LOCUS
DEFINITION
Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
sequence, section 5/9.
ACCESSION
AP003362
VERSION
AP003362.2
KEYWORDS
ORGANISM
Staphylococcus aureus subsp. aureus Mu50
REFERENCE
AUTHORS
Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,
Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Iino, J., Ito, T., Kanamori, M.,
Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A.,
Mizutani, U., Y., Takahashi, N. K., Sawano, T., Inoue, R., Kato, C.,
Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J.,
Kamihisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C.,
Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., and Hiramatsu, K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
JOURNAL
Lancet 357 (9264), 1225-1240 (2001)
MEDLINE
21311952
PUBMED
11418146
REFERENCE
2 (bases 1 to 346900)
AUTHORS
Ohta, T.
TITLE
Direct Submision
JOURNAL
Submitted (28-Feb-2001) Toshiko Ohta, University of Tsukuba College
of Medical Technology and Nursing, Department of Medical
Technology, 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail: tohtaga@tsukuba.ac.jp, Tel: 81-298-3454,
Fax: 81-298-53-3454)
COMMENT
On May 29, 2001 this sequence version replaced gi:13875626.
FEATURES
source
location/Qualifiers
1..346900
/organism="Staphylococcus aureus subsp. aureus Mu50"
/mol_type="genomic DNA"
/strain="Mu50"
/sub_species="aureus Mu50"
/db_xref="taxon:158878"
12..263
/gene="SAV1313"
12..263
/gene="SAV1313"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57475.1"
/db_xref="GI:14247084"
/translacion="MLTLIKLKEDEOVYIIEYIPEDISNGKCSYTFNNKDAEVIDFS
LSEIEEVEFMYLRNKSFSVRDEIEOPEPNKIAMY"
987..1181
/gene="SAV1314"
987..1181
/gene="SAV1314"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57476.1"
/db_xref="GI:14247085"
/translacion="MNLQSYIGIKLVKLTNNKILGKVIDFDQVDNFDGYSIEID
FGRIYDISENKIKIVQK"
complement(1725..2750)
/gene="SAV1315"
complement(1725..2750)
/gene="SAV1315"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57477.1"
/db_xref="GI:14247086"

```

/translatability="MISENDVLEGAHEKYLKRLVDTNINLVQASGCGFPAQIALEI  
 KOTIDEPNATIRLVLGGTNOTNOVYVINSMLSESYSGVSAPTGVAVHEGGALEYGSHVY  
 ITIPSEKGEVSDVETMETFEKSDPFRDHVHPGPNVYISHPTEGTLGKICGLSEIAVE  
 KVKCOPLPLPDGAGRLCYGGLMSQDSOMTIRDAKCDVFGYIGTKICGLSEIAVEY  
 KNEPDPOTFRILKHGALLARGLTGTOTPELTEDLVYNISRAHLEANKMGKGFKN  
 KGRILPDPSPTRNOOFFILSNREKIALEPOKFAVWEKYYDOHRYVRFATSNATTENL  
 NKLEELT  
 gene 3020. .3217  
 /gene="SAV1316"  
 3020. .3217  
 /gene="SAV1316"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAB57478.1"  
 /db\_xref="GI:14247087"  
 /translatability="MEQIKIKTFTTAEVLLELENNINAFSSSEATNLKLVITTKETE  
 ERTPNNEEFNALTLISYVK  
 gene 3238. .4755  
 /gene="SAV1317"  
 3238. .4755  
 /gene="SAV1317"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cardiolipin synthetase homolog"  
 /protein\_id="BAB57479.1"  
 /db\_xref="GI:14247088"  
 /translatability="MEYPPFJGEXTMRKTRPNSDLGTLTITLAIIGFIINLVLAFFFFIT  
 FLENNRSTASTMAFLVFLVPLTIGFLILTFERRVSARKLNKNVNLDTDFGLLK  
 QOULESDPGKQKOVOKHIDVLSMLMDQDGFLENNKVDHFTDSNDYLDVGLK  
 IKNAKEITLHLEYTETPDLGKRLILHALEERLQGLEVLITLDVDSKRYKMANDFH  
 KSLGCEVAEFPAKLPFLNFRMNNNSHAKIIVIDGQGVGGFNIGDEYLDGKIGYVA  
 RDHLRJCQGAVALDQRLREILDMNSQAHKQFEDYDVAFPKNGPLGNSPQIQAASCP  
 ASOMHOEGTEGRTMISAKASYLOSPEYIPDSYINAKRIAKSGVVDHMLPCKEDP  
 HPLVMAETGFSNADLSLSCGVKIYTYENGFIHSKMLIDEIVSGTANMDRSEFLNF  
 EVNAPFYDEVLAKDLAVAEHDIKRSQDLFRESYANRPISVAFKESLAKLVSPIL"  
 gene 4939. .5436  
 /gene="SAV1318"  
 4939. .5436  
 /gene="SAV1318"  
 /note="ATP-binding protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ABC transporter homolog"  
 /protein\_id="BAB57480.1"  
 /db\_xref="GI:14247089"  
 /translatability="MOISINIKRSPFKRCVULKMISEFDIEGCAIALIGNKNGASTLL  
 DILGANNASGEIPKOLKLOSEBRSIMOKTMMPPDQIKVEIINLYOSQENPDL  
 BEIIEILKFPSSQLNPFNKLSCGQRLDFVLSLIGQQLILDEPSTNDIEIREY  
 PMSIIT  
 gene 5461. .5838  
 /gene="SAV1319"  
 5461. .5838  
 /gene="SAV1319"  
 /note="ATP-binding protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ABC transporter homolog"  
 /protein\_id="BAB57481.1"  
 /db\_xref="GI:14247090"  
 /translatability="MITSHTYEYERKSKILIEGELLINDSTSHRNQSOITTE  
 LSDBITKRLKLDLVTOKNHNIGTIKITSNVMDIILYLDQHLINLDDIEIXVSIV  
 DSIFNNKKGQSYDPRKLENNRI  
 gene 5807. .6538  
 /gene="SAV1320"  
 5807. .6538  
 /gene="SAV1320"  
 /note="similar to ABC transporter integral membrane  
 protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAB57482.1"

[illegible]

Db	200423	CGCTGTACGCAATTCGCATATATCTACTTGTTAACGGCTAAATACGACCGGTTTTTAC	2004482
QY	181	ATTGTGTATAAATTTTATGCGACCAATTCATGCGCTTTGTGTCAATCTTACACGACGTTAA	240
Db	200483	ATTGTGTATAAATTTTATGCGACCAATTCATGCGCTTTGTGTCAATCTTACACGACGTTAA	2005424
QY	241	CCATTAATAAATAATACATCCGCTTATCATGATGAGGATTTGCGCGCGGCTCTAATTTGTG	300
Db	200543	CCATTAATAAATAATACATCCGCTTATCATGATGAGGATTTGCGCGCGGCTCTAATTTGTG	200602
QY	301	TTTTGGAATGCTGCATCACCAATTTGCATACCACTGTATCAATAATATGAAATCATGTG	360
Db	200603	TTTTGGAATGCTGCATCACCAATTTGCATACCACTGTATCAATAATATGAAATCATGTG	200662
QY	361	TTTAACCATTCACCTGAAAGAAATAAATAC	387
Db	200663	TTTAACCATTCACCTGAAAGAAATAAATAC	200689

Db	2417	TATCATATACGCATTTCCAAAGATTATTCACCACTTATTATTCACAGCTAATACACAGGTTTCTTAG	2416
Qy	181	ATTGTATATAAATTTGACGACACATTTTCATCGCTTTGTGTGTCAATCCCTTCACGCAGTTAA	240
Db	2477	ATTTATATAAGCATTTGTGTGCGACCAATTCGTCACATTTGTGTAAATCTTCTTCATACATTGA	2536
Qy	241	CCATTAATAATATATACATCGGCTCATCTATGCGATTTCTGTGCTGCTGCTTAATTTGTG	300
Db	2537	CCATTAATAATATGATCATCTGCTTCATCAATGCTATTTCTGTGCTGTGACGAGATTTAG	2596
Qy	301	TTTGGATGTGTCATCACACATTTTCATATACACCTGTATCAATATAATTTGAATCATGTG	360
Db	2597	TTTGAAAAGAGACATCTCCATTTTCATATACACCTGTATCAATGAATGATTTAATTCATAG	2656
Qy	361	TTTACCATTCACCTGACAGATAAATAC	387
Db	2657	TTTACCATTCGCGCAAGATGAATAAATAC	2683

RESULT 5	AF270032	3269 bp	DNA	linear	BCT 01-AUG-2000
LOCUS	Staphylococcus epidermidis strain SRI clone step.1042107				genomic
DEFINITION	sequence.				
ACCESSION	AF270032				
VERSION	AF270032.1	GI:9623936			

RESULT 6			
AX145350			
LOCUS	AX145350	3269 bp	DNA
DEFINITION	Sequence 4072 from Patent WO0134809.		linear
ACCESSION	AX145350		
VERSION	AX145350.1	GI:14283915	
KEYWORDS	.		

SOURCE ORGANISM	REFERENCE
<i>Staphylococcus epidermidis</i>	1 (bases 1 to 369)
<i>Staphylococcus epidermidis</i>	Kimerly, W. J., Taylor, J. David, Nielsen, A. J., Godlewski, M. M.,
Bacteria: Firmicutes: Bacillales: <i>Staphylococcus</i>	

ORGANISM	synthetic construct
REFERENCE	artificial sequences.
AUTHORS	1
TITLE	Kimmerly, W. J. Staphylococcus epidermidis nucleic acids and proteins

TITLE	transposon-mediated sequencing of the <i>Staphylococcus epidermidis</i> genome
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 3269)
AUTHORS	Taylor, J. David., Kimmerly, W. J., Nielsen, A. J., Godlevski, M. M.,

```

source
x:200
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="synthetic nucleic acid sequence"
986 a      451 q 1216 t
616 c

```

TITLE	Direct Submission
JOURNAL	Submitted (22-MAY-2000) Departments of Genomic Sciences and Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES	Location/Qualifiers
source	1. 3269

```

0Y      1 GATCTCTCCCTTACCAAAATGAGAAACACTGACTAACAAGTACCAAGACTA 60
        |||||  |||  |||  |||||  |||||  |||  |||  |||||  |||
Db      2297 GATCTCTGATCTTATTATTAAGTTTCAACAACACTGACTTACAAATCTCCAACTCTA 2356

```

```

/organism="Staphylococcus epidermidis"
/mol_type="genomic DNA"
/strain="SRI"
/db_xref="taxon:1282"
/clone="step.1042f07"
BASE COUNT      986 a      616 c      451 g      1216 t
ORIGIN

```

[illegible]

Query Match	67.8%	Score 262.2	DB 1	Length 3269
Best Local Similarity	79.8%	Pred. No. 3.1e-49		
Matches 309	Conservative	0	Mismatches 78	Indels 0
				Gaps 0

  

Oy	1	GATCTCTCCCTCCACCAAAATGAGAAACAACTGATCTATACAGTACCAAAAGCTTA	60
Db	2297	GATCTCTGATCTCTTATTATTAAGTTTCAACCACTGATCTATACCAAAATCTCAATCTCTTA	2356

[illegible]

OY 61 AACCATGTACCCTATATTCGATACGGTCCACCAATACTCATNGAATAGAAATCATACA 120  
|| |||||||  
Db 2357 ATCCATGTAACCAAAAATAGSAAATGATCTCCAAAAGCTTAAGAATGAATCATATGAA 2416  
OY 121 CGTCTAGCAGATTCCATATTATCTACTTGTTAACCGCTATATACGACCGSTTTTTAG 180  
|| |||||||

[illegible]









gene  
EAKFMGKCDIYKTDREGIT\*  
7854.. .9680  
/gene="pepf-1"

Query Match 54.4%; Score 210.6; DB 1; Length 290117;  
Best Local Similarity 71.7%; Pred No. 9,2e-38;  
Matches 276; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

source

gene

CDS

1..304680  
/organism="Bacillus cereus ATCC 14579"  
/mol\_type="genomic DNA"  
/strain="ATCC 14579"  
/db\_xref="ATCC:14579"  
/db\_xref="taxon:226900"  
68..427  
/locus\_tag="BC1239"  
68..427  
/locus\_tag="BC1239"  
/locus\_tag="BC1239"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical Membrane Spanning Protein"  
/protein\_id="AA08324.1"  
/db\_xref="GI:29894936"  
/transl\_table="MRRLYLFLVPLVANLISMTKIELMALFLGVFISNLRI  
OESILEICFSLSTELSVISVIVLNLFLRHSFTLMKRNHIILOAICITIVFVFL  
KEIVGQIVFQDRKME"  
551..2041  
/locus\_tag="BC1240"  
551..2041  
/locus\_tag="BC1240"  
/locus\_tag="BC1240"  
/codon\_start=1  
/transl\_table=11  
/product="L-lactate permease"  
/protein\_id="AA08325.1"  
/db\_xref="GI:29894937"  
/transl\_table="MLALIPIMHIFICLFLEKQSLNSLSISVSVGIYLSPTOL  
GISETHATIKGLICFLIVGVLFGLFPLHMKNGYIDQVAFLEEVYQSLDLM  
LMCEGPILESVSGFQIPVNAAPISLQKPPQVLAFLGLASSGVAATGTI  
IGSGLIMPLETTLGTNTALLSIPMAFVPLISLHVAGCMQAVIKENEGFLLPSL  
GIYLSNAVSVBELAGILSSIVITFGFLIKKSGCNLTIBHAAATEBEVSIITII  
SPYFLVLCILSRLPALHDLFRSVAVLADLSYSYKLELSYGFNLGCTFTTII  
FRIDSNIKOGLSOTIKQWIPAITTFMFIATSELMGSGHSLATADDTGTFV  
FVAPFICAGFLTGSNAGSNMFIKLOQTONVALPMQVYTLTQWASVATVACP  
SRITIGVLCNIPYRENEELKKTTLITIGAVLVVEVIVVMLELN"  
2132..2818  
/locus\_tag="BC1241"  
2132..2818  
/locus\_tag="BC1241"  
/locus\_tag="BC1241"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA08326.1"  
/db\_xref="GI:29894938"  
/transl\_table="MTKNVGVKGVIOSEIGCKTPHYOVHIOGFTGVDAIATV  
KSGSYPSVLYFRASNNIKSEDAIHLPULPFGTETKNNRPVADYIRGNLPSNOMI  
PLPERKAGADNDLENKTEKIRALBEKAITARQENRGPENTPDSTFRKNGIH  
DIHMOGNEVKMGDDGIWQDGIILHFEKEEEMIGIFLAFQSOSECTDEGHARPV  
EHCDYKCNIX"  
complement(2830..3663)  
/locus\_tag="BC1242"  
complement(2830..3663)  
/locus\_tag="BC1242"  
/codon\_start=1  
/transl\_table=11  
/product="Integral membrane protein"  
/protein\_id="AA08327.1"  
/db\_xref="GI:29894939"  
/transl\_table="MELVNSIFQILTVSLOLSLIGVIVIGFLGYSLTRMYWS  
KAFGRKGFLLTAMIGVPIHEILHAIMCVLFRKIVATOPFTDSOGLAYOHQNO  
KSVORIGNEFIIGIPISIGITALILANKYVPESEYFENLTLEKTIASISINIE  
NMLSTFVLKSLFTIGNLNPSFMLEIAICSAHIALSKPIKSIKSIADGVIVFV  
LLENITTFGLFOYSNQLGOVMKYNMVLIASSVALLFSCISITLSFSFYKILRRGS  
P"  
3860..4630  
/locus\_tag="BC1243"  
3860..4630  
/locus\_tag="BC1243"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA08328.1"

OY 3 TCTTCTCTCTTACCAAAATGAGAAACAAGTGCATCTAAGACGACCAAGCTAA 62  
DB 280879 TCTTCTCTCTTATTTGGAAAAAGCTGTCATCTCATTAAGTCACTTAATCTTAAT 280820  
OY 63 CCATGTACCCCTGATATCGGATACGCTTACCAAAATCTTAATGATGAATATCAGC 122  
DB 280819 CCGTGTACCTGAGATTTGGGAATGCGCGCAAAATCTTAATGATGAATATCAGC 280760  
OY 123 TCTGTACGATTTCCATATATCTATCTTTGTTACCGCTATACGACCGCTTTTATGAT 182  
DB 280759 TCACCTGCGATTTCCGATTTGTCACCTTATTTACTGCACTACACGCTTTATTTGAA 280700  
OY 183 TTGTATAAATTTGACGACCATTTTCATCGCTTGTGTCAATCCCTCAGCAGCTTAAC 242  
DB 280699 CGATATATATTTTTCACACTCTTCATCCGCTGCTGTACACCGTCACGACATTTGTC 280640  
OY 243 ATAAAAATATAATCCGCTCATCTATGGGATTTCTGCGCTGCTAATTTGCT 302  
DB 280639 ATAAAAATATACATCTGCTCATCAATTTGCTACTGCTTGTGACGAAATTTGTGTC 280580  
OY 303 TGAATGTGATCACCACCAATTTCAATACCACCTGTATCAATATATTTGAATCATGTGTT 362  
DB 280579 AAAAAAGCTCGCTCCCAATATCAATTCACCTGTATCAATATTTGAATCATGTGTT 280520  
OY 363 AACCATTCACCTGAGAAATATAATAC 387  
DB 280519 AACCATTCACGACGCTATATAATAC 280495

RESULT 9  
LOCUS AE017002/c 304680 bp DNA linear BCT 16-MAY-2003  
DEFINITION Bacillus cereus ATCC 14579 section 5 of 18 of the complete genome.  
ACCESSION AE017002 AE016877  
VERSION AE017002.1 GI:29894935  
KEYWORDS  
SOURCE Bacillus cereus ATCC 14579  
ORGANISM Bacillus cereus ATCC 14579  
Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus  
cereus group.

REFERENCE 1 (bases 1 to 304680)  
AUTHORS Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelon,B.,  
Kapral,V., Bhattacharya,A., Reznik,G., Mikhailova,N.,  
Lapidus,A., Chu,L., Mazur,M., Goltsman,E., Larsen,N., D'Souza,M.,  
Malunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fomstein,M.,  
Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.  
Genome sequence of Bacillus cereus and comparative analysis with  
Bacillus anthracis  
NATURE 423 (6935), 87-91 (2003)

TITLE  
JOURNAL MEDLINE  
PUBMED 22609415  
13721630  
2 (bases 1 to 304680)  
Candelon,B., Galloux,K., Ehrlich,D.S. and Sorokin,A.  
The number of ribosomal RNA operons in Bacillus cereus  
Unpublished  
3 (bases 1 to 304680)

REFERENCE  
AUTHORS Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelon,B.,  
Kapral,V., Bhattacharya,A., Reznik,G., Mikhailova,N.,  
Lapidus,A., Chu,L., Mazur,M., Goltsman,E., Larsen,N., D'Souza,M.,  
Malunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fomstein,M.,  
Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.  
Direct Submission  
Submitted (12-MAR-2003) INRA, Genetique Microbienne, Domaine de  
Villvert, Jouy en Josas 78352, France  
Location/Qualifiers

TITLE  
JOURNAL  
FEATURES

gene

CDS

gene

CDS

gene

CDS

1..304680  
/organism="Bacillus cereus ATCC 14579"  
/mol\_type="genomic DNA"  
/strain="ATCC 14579"  
/db\_xref="ATCC:14579"  
/db\_xref="taxon:226900"  
68..427  
/locus\_tag="BC1239"  
68..427  
/locus\_tag="BC1239"  
/locus\_tag="BC1239"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical Membrane Spanning Protein"  
/protein\_id="AA08324.1"  
/db\_xref="GI:29894936"  
/transl\_table="MRRLYLFLVPLVANLISMTKIELMALFLGVFISNLRI  
OESILEICFSLSTELSVISVIVLNLFLRHSFTLMKRNHIILOAICITIVFVFL  
KEIVGQIVFQDRKME"  
551..2041  
/locus\_tag="BC1240"  
551..2041  
/locus\_tag="BC1240"  
/locus\_tag="BC1240"  
/codon\_start=1  
/transl\_table=11  
/product="L-lactate permease"  
/protein\_id="AA08325.1"  
/db\_xref="GI:29894937"  
/transl\_table="MLALIPIMHIFICLFLEKQSLNSLSISVSVGIYLSPTOL  
GISETHATIKGLICFLIVGVLFGLFPLHMKNGYIDQVAFLEEVYQSLDLM  
LMCEGPILESVSGFQIPVNAAPISLQKPPQVLAFLGLASSGVAATGTI  
IGSGLIMPLETTLGTNTALLSIPMAFVPLISLHVAGCMQAVIKENEGFLLPSL  
GIYLSNAVSVBELAGILSSIVITFGFLIKKSGCNLTIBHAAATEBEVSIITII  
SPYFLVLCILSRLPALHDLFRSVAVLADLSYSYKLELSYGFNLGCTFTTII  
FRIDSNIKOGLSOTIKQWIPAITTFMFIATSELMGSGHSLATADDTGTFV  
FVAPFICAGFLTGSNAGSNMFIKLOQTONVALPMQVYTLTQWASVATVACP  
SRITIGVLCNIPYRENEELKKTTLITIGAVLVVEVIVVMLELN"  
2132..2818  
/locus\_tag="BC1241"  
2132..2818  
/locus\_tag="BC1241"  
/locus\_tag="BC1241"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA08326.1"  
/db\_xref="GI:29894938"  
/transl\_table="MTKNVGVKGVIOSEIGCKTPHYOVHIOGFTGVDAIATV  
KSGSYPSVLYFRASNNIKSEDAIHLPULPFGTETKNNRPVADYIRGNLPSNOMI  
PLPERKAGADNDLENKTEKIRALBEKAITARQENRGPENTPDSTFRKNGIH  
DIHMOGNEVKMGDDGIWQDGIILHFEKEEEMIGIFLAFQSOSECTDEGHARPV  
EHCDYKCNIX"  
complement(2830..3663)  
/locus\_tag="BC1242"  
complement(2830..3663)  
/locus\_tag="BC1242"  
/codon\_start=1  
/transl\_table=11  
/product="Integral membrane protein"  
/protein\_id="AA08327.1"  
/db\_xref="GI:29894939"  
/transl\_table="MELVNSIFQILTVSLOLSLIGVIVIGFLGYSLTRMYWS  
KAFGRKGFLLTAMIGVPIHEILHAIMCVLFRKIVATOPFTDSOGLAYOHQNO  
KSVORIGNEFIIGIPISIGITALILANKYVPESEYFENLTLEKTIASISINIE  
NMLSTFVLKSLFTIGNLNPSFMLEIAICSAHIALSKPIKSIKSIADGVIVFV  
LLENITTFGLFOYSNQLGOVMKYNMVLIASSVALLFSCISITLSFSFYKILRRGS  
P"  
3860..4630  
/locus\_tag="BC1243"  
3860..4630  
/locus\_tag="BC1243"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA08328.1"









termInator  
5681..5708  
/gene="lmo1694"  
complement(5682..8291,8299..8304))  
/gene="lmo1695"  
complement(5682..5709)  
/gene="lmo1695"  
complement(5694..8291)  
/gene="lmo1695"  
/note="similar to putative membrane proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC99773.1"  
/db\_xref="GI:1641149"  
/db\_xref="SPTREMBL:O8Y619"  
/translacion="MKELMQAYAMFQKNSVVKVIFEFVNAFVIFELINATCIDY  
PSLKENTLOSPEOFIMFIVGLAVTPMLYDVIYVILPKFSPSHVASMINT  
FTNIGFCGVLGASRASFGKNSHKREITLAIKALFVGLSTIYCLVSLATLIP  
GFAHPVNYMPLAGLIPPLFTITRKMSKSLFVDLPITKRELIITLISLEMGRF  
GFAITGLMEKSPDIFKVPFLPIYASVIGTASVPGCTPPVATILASOLVSOE  
LALAMLTIRITITIPVGLTFVQKQKNDLEGLPLFLQKVAHFLVTV  
GSLILLISSAIFNTLAVFSLMKAIFGLVSLQITLAFNEFVREKLVYTSKVIID  
AYITIVVGAIFNTLAVFSLMKAIFGLVSLQITLAFNEFVREKLVYTSKVIID  
SITFVCLAGVIGVIVNSPNIKHSEIPYLRIASEHMLVGVFAVSVLVIY  
IYLSTKEKLSPEAVNSPEHRLAKMGNEVSHMFLRDLKFLMAEGELFSYRILA  
DKMVLNGEPTGMOMKMEALIEVYMNADREFYRVEYVAGTMIPLYLHDGDFETIA  
EDGFVONFTMSGKKKGERALANKLERGTYFEIIEPENNIDTTLTILAVSDKLD  
GREGKFSIGFEDFDYLEOAPITAKNGDETIYFASMSYSDTDEMTSIDLRYSSEA  
PSGIMDPLPIPLPEKAKEDGQTFNAGKMLAVNGSEKVFLEERLIVYVSCYGF  
GPKGRNRSKYVTEDQKPVAFPRKRSIAFTHLQIMLIVGKKRPLANSQVLDPELE  
ETKKKDESE

RBS  
complement(8299..8304)  
/gene="lmo1695"  
complement(8318..9007)  
/gene="lmo1696"  
complement(8318..8992)  
/gene="lmo1696"

Query Match 53.2% Score 205.8: DB 1: Length 347050;

Best Local Similarity 70.9%: Pred. No. 1:1e-36; Mismatches 112: Indels 0: Gaps 0:

Matches 273: Conservative 0: Mismatches 112: Indels 0: Gaps 0:

QY 3 TCTCTCTCTTACCAAAATGAGAAACATCTCATCTAACAAGTCCAGCAAGCTTAA 62  
DB 258933 TCCTCTCTCTTGGAAAATGAGCCGCAACACATCAAGCTGCGCAACCCCTAGT 258992  
QY 63 CCATGAGCCCTGATATGAGATACGGTTCACCAATCCTATGAAATGAAATCATACAG 122  
DB 258993 CCATGAGCAAGAAATGAGATACGGTTCACCAACCAAGAAATGAAATCATAAAT 259052  
QY 123 TCTGTACGATTCATATCTACTTGTGTAACCGCTATACGACCGCTTTTAAAT 182  
DB 259053 TGATCTGCAATTCGCGTTATCTATTAATCCCTAAACAAATGGTTATTAAGAC 259112  
QY 183 TTGTATAAATTTGAGCGACATTCATCGCTTGTGTCAATCCCTACGACAGCTTAAC 242  
DB 259113 CGGTAAAGAAATTTGTAACCTGTCGTACGTAACCTCCTACGACAGCTTGTGA 259172  
QY 243 ATAAAAATATAAATCGGCTTCATCTATGGCGATTCGCGTCCGCTGCTAAATTTGTT 302  
DB 259173 ATAAAAATATAAATCGGCTTCATCTATGGCGATTCGCGTCCGCTGCTAAATTTGTT 259232  
QY 303 TGGATGTGATACCAATTTCAATACCACTGTATCAATATATTAATGAAATCATGTGTT 362  
DB 259233 AAAAAATGATTCGAAAGATCAATACCACTGTATCAATATATTAATGAAATTTCTTTCCA 259292  
QY 363 AACCATTCACCTGAGAAATTAATAC 387  
DB 259293 AGCCATTCGCGTGAATTAATATATGC 259317

RESULT 13

AX417046  
LOCUS AX417046 349980 bp DNA linear PAT 02-SEP-2002  
DEFINITION Sequence 4037 from Patent WO0228891.  
ACCESSION AX417046  
VERSION AX417046.1 GI:21449656  
KEYWORDS  
SOURCE  
ORGANISM  
1. *Listeria innocua*  
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
REFERENCE  
1. Kunst, F. and Glaser, P.  
AUTHORS  
TITLE *Listeria innocua*, genome and applications  
JOURNAL PATENT: WO 0228891-A 4037 11-APR-2002;  
INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE  
SCIENTIFIQUE (CNRS) (FR)  
FEATURES  
SOURCE  
Location/Qualifiers  
1. 349980  
/organism="Listeria innocua"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1642"  
/note="seq 2058, original length: 3,011,208 replaced  
by-seq 2058: 0.000.001 to 0.349.980-seq 4032: 0.300.001 to  
0.649.980-seq 4033: 0.600.001 to 0.949.980-seq 4034:  
0.900.001 to 1.249.980-seq 4035: 1.200.001 to  
1.549.980-seq 4036: 1.500.001 to 1.849.980-seq 4037:  
1.800.001 to 2.149.980-seq 4038: 2.100.001 to  
2.449.980-seq 4039: 2.400.001 to 2.749.980-seq 4040:  
2.700.001 to 3.049.980-seq 4041: 3.000.001 to 3.011.208"  
ORIGIN  
101055 a 72969 c 60688 g 115268 t

Query Match 53.2% Score 205.8: DB 6: Length 349980;

Best Local Similarity 70.9%: Pred. No. 1:1e-36; Mismatches 112: Indels 0: Gaps 0:

Matches 273: Conservative 0: Mismatches 112: Indels 0: Gaps 0:

QY 3 TCTCTCTCTTACCAAAATGAGAAACATCTCATCTAACAAGTCCAGCAAGCTTAA 62  
DB 26305 TCCTCTCTCTTGGAAAATGAGCCGCAACACATCAAGCTGCGCAACCCCTAGT 26364  
QY 63 CCATGAGCCCTGATATGAGATACGGTTCACCAATCCTATGAAATGAAATCATACAG 122  
DB 26365 CCATGAGCAAGAAATGAGATACGGTTCACCAACCAAGAAATGAAATCATAAAT 26634  
QY 123 TCTGTACGATTCATATCTACTTGTGTAACCGCTATACGACCGCTTTTAAAT 182  
DB 266425 TGATCTGCAATTCGCGTTATCTATTAATCCCTAAACAAATGGTTATTAAGAC 26684  
QY 183 TTGTATAAATTTGAGCGACATTCATCGCTTGTGTCAATCCCTACGACAGCTTAAC 242  
DB 266485 CGGTAAAGAAATTTGTAACCTGTCGTACGTAACCTCCTACGACAGCTTGTGA 26544  
QY 243 ATAAAAATATAAATCGGCTTCATCTATGGCGATTCGCGTCCGCTGCTAAATTTGTT 302  
DB 26545 ATAAAAATATAAATCGGCTTCATCTATGGCGATTCGCGTCCGCTGCTAAATTTGTT 26604  
QY 303 TGGATGTGATACCAATTTCAATACCACTGTATCAATATATTAATGAAATCATGTGTT 362  
DB 26605 AAAAAATGATTCGAAAGATCAATACCACTGTATCAATATATTAATGAAATTTCTTTCCA 26664  
QY 363 AACCATTCACCTGAGAAATTAATAC 387  
DB 26665 AGCCATTCGCGTGAATTAATATATGC 266689

RESULT 14  
AX641670  
LOCUS AX641670 349980 bp DNA linear PAT 21-FEB-2003  
DEFINITION Sequence 2860 from Patent WO0101118.  
ACCESSION AX641670  
VERSION AX641670.1 GI:28474431  
KEYWORDS  
SOURCE  
ORGANISM  
1. *Listeria monocytogenes*



```

REFERENCE 1 Bacteria; Firmicutes; Bacilliales; Listeriaceae; Listeria.
AUTHORS Buchrieser, C., Frangeul, L., Couve, E., Rusniok, C., Fsihi, H.,
Dehoux, P., Dussurget, O., Chetoui, F., Nedjari, H., Glaser, P.,
Kunst, F., Cossart, P., Daniels, J., Goebel, M., Kreft, J., Kuhn, M.,
Ng, E., Vasquez-Boland, J., Dominguez-Bernal, G., Garrido-Garcia, P.,
Tlerraz-Martinez, A., Amend, A., Charabory, T., Domann, E., Halin, T.,
Berche, P., Charbit, A., Durant, L., Perez-Diaz, J. C., Bagnero, F.,
garcia del Portillo, F., Gomez-Lopez, M., Madueno, E., de Pablo, B.,
Weiland, J., Kaeber, U., Entian, K. D., Hauf, J., Rose, H., and Voss, H.,
Listeria monocytogenes genome, polypeptides and uses
JOURNAL Patent: WO 010118-A 2860 11-APR-2001;
INSTITUT PASTEUR (FR)
TITLE Location/Qualifiers
JOURNAL 1. 349980
FEATURES
source
/organism="Listeria monocytogenes"
/mol_type="genomic DNA"
/db_xref="taxon:1639"
/notes="Seq 1 for big: original length: 2944528-bp, split
into 10 <23>seq 0001 : From 0.000.001 to 0.349.980
length: 349.980 <23>seq 2853 : From 0.300.001 to
0.649.980 length: 349.980 <23>seq 2856 : From 0.600.001
to 0.949.980 length: 349.980 <23>seq 2857 : From
0.900.001 to 1.249.980 length: 349.980 <23>seq 2858 :
From 1.200.001 to 1.549.980 length: 349.980 <23>seq 2859
: From 1.500.001 to 1.849.980 length: 349.980 <23>seq
2860 : From 1.800.001 to 2.149.980 length: 349.980
<23>seq 2861 : From 2.100.001 to 2.449.980 length:
349.980 <23>seq 2862 : From 2.400.001 to 2.749.980
length: 349.980 <23>seq 2863 : From 2.700.001 to
2.944.528 length: 244.528"
BASE COUNT 99761 a 75437 c 61177 g 113605 t
ORIGIN

Query Match 53.24; Score 205.8; DB 6; Length 349980;
Best Local Similarity 70.9%; Pred. No. 1,1e-36;
Matches 273; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 3 TCTTCTTCCTTCACCAAAATGAGAAACATCGCATCTAACAAGTCAACCAAGCTTAA 62
DB 211933 TCCCTCTCTCTTTTGGAAATATGACGCCGACACATCAACGATCGCCATGCT 211992
QY 63 CCATGTGACCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 122
DB 211993 CCATGTGACCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 212052
QY 123 TCTGTACGATTCATATATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 182
DB 212053 TGATCTGCAATTCGTGCTATCTATCTATCTATCTATCTATCTATCTATCTAT 212112
QY 183 TTGTATTAATTTGAGGACGACATTCATCGCTTGTGTCAATCTTCACGACGTTAAC 242
DB 212113 CGGTAAAGAAATTTTCTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 212172
QY 243 ATAAATAATTAATACATCGCTTCATCTATGCGGATTTTCGCTGCGCTCTATTTGTGT 302
DB 212173 ATAAATAATTAATACATCGCTTCATCTATGCGGATTTTCGCTGCGGATTTGTCT 212232
QY 303 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 362
DB 212233 AAAAATGCTTCATCGGAAATATCAATCAACCGCTGATTCGCTGCGGATTTGTCT 212292
QY 363 AACCATTCACCTGAAAGATTAATAC 387
DB 212293 AGCCATTCGCCCTGAATTAATATAC 212317

RESULT 15
BACSERA 24887 bp DNA linear BCF 09-DEC-1998
LOCUS BACSERA/c
DEFINITION Bacillus subtilis phosphoglycerate dehydrogenase (serA), ypaA,
ferredoxin (fer), ypbB, recS, ypbD, ypbE, ypbF, ypbG, ypbH,
glutamate dehydrogenase (ypcA), ypdA, ypbB, ypbC, spore cortex

```

lytic enzyme (slsB), ypeB, ypfA, ypfB, cyclidine monophosphate kinase (cmk), ypiD, ypgA, yphA, yphB, yphC, NAD+ dependent glycyl-3-phosphate dehydrogenase (glyC), yphE and yphF genes, complete cds.

L47648.1 GI:1146195

Bacillus subtilis  
Bacillus subtilis  
Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.

1 (bases 24112 to 24887)

Roels, S., Driks, A. and Losick, R.

Characterization of spoIVA, a sporulation gene involved in coat morphogenesis in Bacillus subtilis.

J. Bacteriol. 174 (2), 575-585 (1992)

92105027

1729246

2 (bases 1 to 24887)

Sorokin, A., Zumstein, E., Azevedo, V., Ehrlich, S.D. and Serror, P.

The organization of the Bacillus subtilis 168 chromosome region between the spoVA and spoVA genetic loci, based on sequence data

Mol. Microbiol. 10 (2), 385-395 (1993)

95020538

7934829

3 (bases 1 to 24887)

Sorokin, A., Azevedo, V., Zumstein, E., Galleron, N., Ehrlich, S.D. and Serror, P.

Sequence analysis of the Bacillus subtilis chromosome region between the spoVA and spoVA loci cloned in a yeast artificial

chromosome

Microbiology 142 (Pt 8), 2005-2016 (1996)

96349105

8760912

Location/Qualifiers

1. 24887

/organism="Bacillus subtilis"

/mol\_type="genomic DNA"

/strain="Marburg"

/db\_xref="taxon:1423"

/complement(1..41)

/citation(2)

/evidence="experimental"

/complement(40..1617)

/gene="serA"

/complement(40..1617)

/gene="serA"

/function="serine biosynthesis"

/citation(12)

/citation(13)

/evidence="experimental"

/transl\_table=1

/product="phosphoglycerate dehydrogenase"

/protein\_id="AAC83943.1"

/db\_xref="GI:1146195"

/translation="MFRVLYSDKMSNDGLOPILESDFEIVQXNVADEDELPHFDAL  
LRSATVTEDELFRKMSLKYGRACVGDNDIDQATDHGCIYVINAQNTISTFAR  
TFAMSSLMRIEIPANISVSKRENNRATAYGSLYKTLGILGRLISEYIAORRAF  
GMYHVPDFLIERANKIGVNSRTEFEVLESADITVHTPLKTRKCLNKKETIAT  
KGVRLINCARGGIIIDEALLLEALENGHAGALDVEFEPVNDKLVDPVLATPH  
LGASTREANQVNAQVSEVLAQFAGKLPMVSAIINLPMATKDEFARKIPYHQIACKS  
LVSOCKMEPVQDVAIQYEGYIAKLLETSTFKALISGFLKPRDVSIVNEVNAQVAKSR  
GISFSEKISSSESGYDNCISVKTGDSFTYATYIPIHGERIYEINGFNIDYPTG  
HUYVIOHODTGVIGVGRITLGDNDININAMQVGRKKEGEAIMLSFDRHLEDKIVK  
ELTNVPOIVSKLIDLP"

2114..2686

/gene="ypaA"

2114..2686

/gene="ypaA"

/note="putative"

/citation=[3]



Sat Sep 13 12:47:59 2003

us-09-815-242-1463.rge

Page 19

Db 21432 CGCTTCATCCATGGCGATTTCAGCTTGCCTGGCGCAATCTGCGCTAAAAACGGCTCATCACC 21373  
Oy 320 AATTTCAATACCACTGTATCATTAATTAATTGAATCATGTGTTAAACATTACACTGAGAGA 379  
Db 21372 GATATCAATACCGCGCGGTATCATCAATAATTAATTAATTCATATTCAGGCATTCAGCCGAGCT 21313  
Oy 380 ATAAATAC 387  
Db 21312 GTATATCC 21305

Search completed: September 12, 2003, 19:27:28  
Job time : 1589.02 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 17:16:39 ; Search time 438.085 Seconds  
(without alignments)  
8041.284 Million cell updates/sec

Title: US-09-815-242-4228

Perfect score: 1305

Sequence: 1 atgactaacctatagtc.....atattatagctcgaagaga 1305

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq.19Jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*
- 25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1305	100.0	1305	23 AAS51646	Staphylococcus aur
2	1305	100.0	1311	23 AAS54865	Staphylococcus aur
3	1292.2	99.0	1311	22 AAF8461	Staphylococcus aur
4	1290.6	98.9	1311	23 AAS54997	Staphylococcus aur
5	1230.2	94.3	3621	18 AAV74669	Staphylococcus epi
6	921	70.5	1312	24 ABR00883	S. epidermidis gen
7	919.4	70.5	3269	22 AAF54708	Listeria monocytog
8	683.2	52.4	2944528	24 ABA03041	

c	9	676.8	51.9	319630	24	ABQ671194	Listeria innocua c
c	10	676.8	51.9	3011208	24	ABQ69245	S. epidermidis ope
c	11	667.6	51.2	960	22	AAH53683	Streptococcus poly
c	12	600	46.0	1308	24	ABN68457	Bacillus lichenifo
c	13	600	46.0	2155561	24	ABN71527	Genomic sequence o
c	14	599.4	45.9	1311	24	ABK75008	S. pneumoniae type
c	15	582.4	44.6	2365589	24	ABA90521	ypc gene of strep
c	16	550.4	42.2	1308	25	ABX07474	Streptococcus pneu
c	17	550.4	42.2	1311	21	AAAS4516	Streptococcus pneu
c	18	550.4	42.2	1311	21	AAAS4516	Streptococcus pneu
c	19	550.4	42.2	5066	19	AAV52212	Streptococcus pneu
c	20	550.4	42.2	2162598	25	ABS56454	Streptococcus pneu
c	21	549.6	42.1	1308	24	ABN68458	Streptococcus pneu
c	22	383.8	29.4	387	23	AAAS4886	Streptococcus pneu
c	23	367.2	28.1	372	23	AAAS0706	Streptococcus pneu
c	24	336.8	25.8	2791	24	ABO70805	Listeria monocytog
c	25	318.4	24.4	738	23	AAAS5783	Listeria monocytog
c	26	308.6	23.6	675	23	AAAS1384	Streptococcus pneu
c	27	308.6	23.6	678	23	AAAS2955	Enterococcus faeca
c	28	296.4	22.7	298	23	AAAS0205	Enterococcus faeca
c	29	296.4	22.7	298	23	AAAS0723	Staphylococcus aur
c	30	293.6	22.5	580073	18	AAAT58840	Staphylococcus aur
c	31	270.8	20.8	33140	22	AAF28536	Mycoplasma genital
c	32	241.8	18.5	640681	24	ABA92787	Genomic fragment #
c	33	241.2	18.5	246	23	AAAS49268	Buchnera sp. genom
c	34	210.4	16.1	1512	22	AAF94379	Haemophilus influe
c	35	210.4	16.1	1512	24	ABK64943	Haemophilus influe
c	36	210.4	16.1	1515	23	AAAS3235	DNA encoding Heemo
c	37	208.8	15.0	1830121	17	AAAT42063	Haemophilus influe
c	38	203.4	15.6	78845	21	AAAB1463	Haemophilus influe
c	39	203.4	15.6	349980	21	AAF21608	N. meningitidis pa
c	40	203.4	15.6	1437668	21	AAAB1490	N. meningitidis B
c	41	198.6	15.2	910715	20	AAAT20248	Borrelia burgdorfe
c	42	190.6	14.6	1500	25	ABZ39377	N. gonorrhoeae nuc
c	43	190	14.6	1512	23	AAAS2517	E. coli DNA for ce
c	44	190	14.6	11574	22	AAAS46244	DNA encoding novel
c	45	178.6	13.7	1557	22	AAH66542	C glutamicum codin

ALIGNMENTS

RESULT 1

AAS51646

ID AAS51646 standard; DNA: 1305 BP.

XX AC AAS51646;

XX AC AAS51646;

DT 13-FEB-2002 (first entry)

XX DE Staphylococcus aureus DNA for cellular proliferation protein #63.

XX XX Antisense; ds: prokaryotic cellular proliferation gene;

XX KW antibiotic; antibacterial; drug design.

XX OS Staphylococcus aureus.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-207727P.

XX PR 23-OCT-2000; 2000US-242578P.

XX PR 27-NOV-2000; 2000US-253625P.

XX PR 22-DEC-2000; 2000US-257931P.

XX PR 16-FEB-2001; 2001US-269308P.

XX XX (ELIT-) ELITRA PHARM INC.

XX XX

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI: 2001-611495/70.  
 DR P-PSDB: AAU33787.  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Claim 27: Seq ID No 4228; 51pp; English.  
 XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding, to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1305 BP; 447 A; 181 C; 277 G; 400 T; 0 other;

Query Match 100.0%; Score 1305; DB 23; Length 1305;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-290;  
 Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGACTAAACCTATAGTACCTATTTAGTGTAGGTAGGCTTAATGTAGGTAAATCTACAATTTT 60  
 DB |||||||  
 OY 1 ATGACTAAACCTATAGTACCTATTTAGTGTAGGTAGGCTTAATGTAGGTAAATCTACAATTTT 60  
 DB |||||||  
 OY 61 AATAGAATAGTTGGAGAACGTTGTTTCGATTTGTTGAAGATACGCCAGGTGTAACACGAGAT 120  
 DB |||||||  
 OY 61 AATAGAATAGTTGGAGAACGTTGTTTCGATTTGTTGAAGATACGCCAGGTGTAACACGAGAT 120  
 DB |||||||  
 OY 121 CGTATTTATCTTCAGGTGAGTGTAAACACATGATTTCAATATATTATGATACAGTGGT 180  
 DB |||||||  
 OY 121 CGTATTTATCTTCAGGTGAGTGTAAACACATGATTTCAATATATTATGATACAGTGGT 180  
 DB |||||||  
 OY 181 ATGAAATAGTTGATGACCATTCACAAACAAATTAGAGCCGACGAGAAATCCGCATA 240  
 DB |||||||  
 OY 181 ATGAAATAGTTGATGACCATTCACAAACAAATTAGAGCCGACGAGAAATCCGCATA 240  
 DB |||||||  
 OY 241 GATGAAGCGGATGTTATTTATTTATGTTTAACTGCGTGAAGGATTGACACAAAGCGAT 300  
 DB |||||||  
 OY 241 GATGAAGCGGATGTTATTTATTTATGTTTAACTGCGTGAAGGATTGACACAAAGCGAT 300  
 DB |||||||  
 OY 301 GAATGGTTCGCTCAATTTTATACAAATCTAAAAACCGGTGCTATTAGCGGTAAACAAA 360  
 DB |||||||  
 OY 301 GAATGGTTCGCTCAATTTTATACAAATCTAAAAACCGGTGCTATTAGCGGTAAACAAA 360  
 DB |||||||  
 OY 361 GTAGATAATATGGAATTCGCTACAGACGCTGATGATTTCTATTATTAGGATTTGCGTAA 420  
 DB |||||||  
 OY 361 GTAGATAATATGGAATTCGCTACAGACGCTGATGATTTCTATTATTAGGATTTGCGTAA 420  
 DB |||||||  
 OY 421 CCGTATCAATATCAGGTGATGATTTAGGCTTGGTGGTCTGTTAGATCCATGTTT 480  
 DB |||||||  
 OY 421 CCGTATCAATATCAGGTGATGATTTAGGCTTGGTGGTCTGTTAGATCCATGTTT 480  
 DB |||||||  
 OY 481 TCTCATTTTGGTGAAGAGGAGAGATTCCTTTATGATGAAGATACAAATTCGACTATCCATT 540  
 DB |||||||  
 OY 481 TCTCATTTTGGTGAAGAGGAGAGATTCCTTTATGATGAAGATACAAATTCGACTATCCATT 540  
 DB |||||||

OY 541 ATTTGACGACCAACAGCTAGGTAAATCAAGTTTAAATGCTATTTAGGTGAAGATCCG 600  
 DB |||||||  
 OY 541 ATTTGACGACCAACAGCTAGGTAAATCAAGTTTAAATGCTATTTAGGTGAAGATCCG 600  
 DB |||||||  
 OY 601 GTTATCGTTTCTAAATGTTTGCAGGACCAACAGAGACGCTATTATGATACAGAGTATATTTAT 660  
 DB |||||||  
 OY 601 GTTATCGTTTCTAAATGTTTGCAGGACCAACAGAGACGCTATTATGATACAGAGTATATTTAT 660  
 DB |||||||  
 OY 661 GATGGACAAGATTTATGTTTAAATGCTGCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
 DB |||||||  
 OY 661 GATGGACAAGATTTATGTTTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
 DB |||||||  
 OY 721 GAATCAACTGAGAAATATTCAGTATTAAAGAGCTTTAAAGAGGATTTAAAGGATTTAAAGGATTT 780  
 DB |||||||  
 OY 721 GAATCAACTGAGAAATATTCAGTATTAAAGAGCTTTAAAGAGGATTTAAAGGATTTAAAGGATTT 780  
 DB |||||||  
 OY 781 GTTTTATGCTGTTATTTGATGCAAGCAAGGCAATTTATGCAACAGATAAAGCGTTGCGAGA 840  
 DB |||||||  
 OY 781 GTTTTATGCTGTTATTTGATGCAAGCAAGGCAATTTATGCAACAGATAAAGCGTTGCGAGA 840  
 DB |||||||  
 OY 841 TATGCACATGAACAAGGTAAAGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 DB |||||||  
 OY 841 TATGCACATGAACAAGGTAAAGCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 DB |||||||  
 OY 901 AAAGATAGTAAACAGATGAAGAAATTTGCAAGATCAAGTACGTAAAGCAATTTCCAAATTTTAA 960  
 DB |||||||  
 OY 901 AAAGATAGTAAACAGATGAAGAAATTTGCAAGATCAAGTACGTAAAGCAATTTCCAAATTTTAA 960  
 DB |||||||  
 OY 961 GATTATGCAACAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 DB |||||||  
 OY 961 GATTATGCAACAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 DB |||||||  
 OY 1021 CCTTACATTAATGAAGCAAGTGAAGCAATTTCCATGAAACCCATACCAACAGCAAGATTTAGCTTAAAT 1080  
 DB |||||||  
 OY 1021 CCTTACATTAATGAAGCAAGTGAAGCAATTTCCATGAAACCCATACCAACAGCAAGATTTAGCTTAAAT 1080  
 DB |||||||  
 OY 1081 GAAGTCTTACTGATGCAATTTCCATGAAACCCATACCAACAGCAAGATTTAGCTTAAATTTTAA 1140  
 DB |||||||  
 OY 1081 GAAGTCTTACTGATGCAATTTCCATGAAACCCATACCAACAGCAAGATTTAGCTTAAATTTTAA 1140  
 DB |||||||  
 OY 1141 AATGCTTTTATGCAACCAAGTTGCTATAGAACCCAGCATTTGCTGTTGTTGTTGTTGTTGTTGTTGTT 1200  
 DB |||||||  
 OY 1141 AATGCTTTTATGCAACCAAGTTGCTATAGAACCCAGCATTTGCTGTTGTTGTTGTTGTTGTTGTTGTT 1200  
 DB |||||||  
 OY 1201 GATGTAGATTAATGCATTTTCTTATAAAGCGTATTAGAGATCAAAATCCGTCGCGCT 1260  
 DB |||||||  
 OY 1201 GATGTAGATTAATGCATTTTCTTATAAAGCGTATTAGAGATCAAAATCCGTCGCGCT 1260  
 DB |||||||  
 OY 1261 TTTGGTTTTGAAAGGTACACCAATTCATATTATAGCTCGAAAGAGA 1305  
 DB |||||||  
 OY 1261 TTTGGTTTTGAAAGGTACACCAATTCATATTATAGCTCGAAAGAGA 1305  
 DB |||||||

RESULT 2

AAS4865  
 ID AAS4865 standard; DNA; 1311 BP.

XX AAS4865;

XX 13-FEB-2002 (first entry)

XX Staphylococcus aureus DNA for cellular proliferation protein #1177.

XX Antisense; ds: prokaryotic cellular proliferation gene;

XX antibiotic; antibacterial; drug design.

XX Staphylococcus aureus.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207272P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253645P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI: 2001-611495/70.  
 DR P-PSDB: AAU37006.  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 PS Claim 27: Seq ID No 8502; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1311 BP; 451 A; 181 C; 277 G; 402 T; 0 other;  
 Query Match 100.0%; Score 1305; DB 23; Length 1311;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-290;  
 Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGACTAAACCTATAGTAGCTATTGTAGGTAGGCGCTAATGTAGGTAATCTACAAATTTT 60  
 DB 1 ATGACTAAACCTATAGTAGCTATTGTAGGTAGGCGCTAATGTAGGTAATCTACAAATTTT 60  
 QY 61 AATAGATAGTGGAGAACGTGTTTCGATTGTTGAAGATACCCAGGTGTACACAGCAT 120  
 DB 61 AATAGATAGTGGAGAACGTGTTTCGATTGTTGAAGATACCCAGGTGTACACAGCAT 120  
 QY 121 CGTATTTATTTCTTCAGGTAGTGGTTAAACACATGATTTCATATTATTGATACAGGTGGT 180  
 DB 121 CGTATTTATTTCTTCAGGTAGTGGTTAAACACATGATTTCATATTATTGATACAGGTGGT 180  
 QY 181 ATTCAAATTTGGTATGTCACCATTCGCAACACAAATTAGAGCCGACGCAATGCCCAT 240  
 DB 181 ATTCAAATTTGGTATGTCACCATTCGCAACACAAATTAGAGCCGACGCAATGCCCAT 240  
 QY 241 GATGAACGGATGTTATTATTTATGTTAAGTCGCGTGAAGGATTGACACAAAGCGAT 300  
 DB 241 GATGAACGGATGTTATTATTTATGTTAAGTCGCGTGAAGGATTGACACAAAGCGAT 300  
 QY 301 GAAATGGTGGCTCAAAATTTTATACAAATCTAAAAACCGGTGCTATTAGCGGTTAACAAA 360  
 DB 301 GAAATGGTGGCTCAAAATTTTATACAAATCTAAAAACCGGTGCTATTAGCGGTTAACAAA 360  
 QY 361 GTAGATAATATGGAATGCGGTACAGACGTGTATGATTTCTATTATAGGATTTGGTGAA 420

DB 361 GTAGATAATATGGAATGCGGTACAGACGTGTATGATTTCTATTCAATAGGATTTGGTGAA 420  
 QY 421 CCGTATCCAATATCAGGGTCACATGTTTGGTCTAGTCTTGGTACATGTTGTTAGATGAGTGT 480  
 DB 421 CCGTATCCAATATCAGGGTCACATGTTTGGTCTAGTCTTGGTACATGTTGTTAGATGAGTGT 480  
 QY 481 TCTCATTTTGGTGAAGAGGAAGAGATCCTTATGATGAAGATACAAATTCGACTATCCATT 540  
 DB 481 TCTCATTTTGGTGAAGAGGAAGAGATCCTTATGATGAAGATACAAATTCGACTATCCATT 540  
 QY 541 ATTGCACACCAACAGGTAGTAAATCAAGTTTAAATGCTTAAATGCTTAAATGCTTAAATGCT 600  
 DB 541 ATTGCACACCAACAGGTAGTAAATCAAGTTTAAATGCTTAAATGCTTAAATGCTTAAATGCT 600  
 QY 601 GTTATCGTTTCTTAATGTTGAGGACCAACGAGAGACGCTATTGATACAGAGTATAGTTAT 660  
 DB 601 GTTATCGTTTCTTAATGTTGAGGACCAACGAGAGACGCTATTGATACAGAGTATAGTTAT 660  
 QY 661 GATGACAAGATTTATGTTTAAATCGATCTGCTGGTATGCTTAAATGCTTAAATGCTTAAATGCT 720  
 DB 661 GATGACAAGATTTATGTTTAAATCGATCTGCTGGTATGCTTAAATGCTTAAATGCTTAAATGCT 720  
 QY 721 GAATCACTGAGAAATATTCAGTATTAAAGAGCTTTAAAGCGATTGAACGTTCAAAATGTT 780  
 DB 721 GAATCACTGAGAAATATTCAGTATTAAAGAGCTTTAAAGCGATTGAACGTTCAAAATGTT 780  
 QY 781 GTTTTAGTGGTTATTGATGCAACAAGAGGATTTATGCAACAGATTAACGTTGTCAGGA 840  
 DB 781 GTTTTAGTGGTTATTGATGCAACAAGAGGATTTATGCAACAGATTAACGTTGTCAGGA 840  
 QY 841 TATGCACATGAACAGGTAAAGCAGTCTGATGTTGCTGTAATTAATGCGATGATCTGTTGAA 900  
 DB 841 TATGCACATGAACAGGTAAAGCAGTCTGATGTTGCTGTAATTAATGCGATGATCTGTTGAA 900  
 QY 901 AAAGATAGTAAACAGCATGAAGAAATTTTGAAGATGAAGTACGTAAAGAAATTTCAATTTT 960  
 DB 901 AAAGATAGTAAACAGCATGAAGAAATTTTGAAGATGAAGTACGTAAAGAAATTTCAATTTT 960  
 QY 961 GATTATGCACAAATTTGCTTTGCTGCTAAAGAGGCAACAGATTACGTACATTTATTC 1020  
 DB 961 GATTATGCACAAATTTGCTTTGCTGCTAAAGAGGCAACAGATTACGTACATTTATTC 1020  
 QY 1021 CTTTACATTAATGAAGCAAGTGAACCATTAACCAAGTCTTCAAAAGTTCAACTTTAAAT 1080  
 DB 1021 CTTTACATTAATGAAGCAAGTGAACCATTAACCAAGTCTTCAAAAGTTCAACTTTAAAT 1080  
 QY 1081 GAAGTCGTTACTGATGCAATTTCCATGCAACCCCTACACCAACAGGTAGAGGTTG 1140  
 DB 1081 GAAGTCGTTACTGATGCAATTTCCATGCAACCCCTACACCAACAGGTAGAGGTTG 1140  
 QY 1141 AATGCTTTTATGCAACCAAGTTGCTTATAGAACCCGACATTTGTTGTTTATTTGTTAAAT 1200  
 DB 1141 AATGCTTTTATGCAACCAAGTTGCTTATAGAACCCGACATTTGTTGTTTATTTGTTAAAT 1200  
 QY 1201 GATGTAGAAATTAATGCAATTTTCTTATTAACCGCTATTAGAGAAATCAAAATCCGCCCT 1260  
 DB 1201 GATGTAGAAATTAATGCAATTTTCTTATTAACCGCTATTAGAGAAATCAAAATCCGCCCT 1260  
 QY 1261 TTTGGTTTTGAGGTACACCAATTCATATTATAGTCTGGAAGAGA 1305  
 DB 1261 TTTGGTTTTGAGGTACACCAATTCATATTATAGTCTGGAAGAGA 1305

RESULT 3  
 AAF86461  
 ID AAF86461 standard; DNA: 1311 BP.  
 XX  
 AC AAF86461;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Staphylococcus aureus YphC coding sequence.

XX YphC; antimicrobial; cytostatic; antiulcer; microbial infection;  
 KW gene therapy; vaccine; gastrointestinal carcinoma; gastric ulcer;  
 KW gastritis; ds.  
 XX Staphylococcus aureus.  
 XX  
 FH Key Location/Qualifiers  
 FT 1..1311  
 FT /tag- a  
 FT /product- "Staphylococcus aureus yphC protein"  
 XX WO200123418-A1.  
 XX 05-APR-2001.  
 XX 19-SEP-2000; 2000WO-US25566.  
 XX 28-SEP-1999; 99US-040968.  
 XX (SHIK ) SMITHKLINE BEECHAM CORP.  
 PA (SHIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Zalacain M, Biswas S, Burnham MKR, Sylvester D, Mcdevitt D;  
 PI Mathie TB;  
 XX WPI: 2001-308138/32.  
 DR P-PSDB; AAB82089.  
 XX  
 PT Novel yphC polypeptides of Staphylococcus aureus useful for diagnosing  
 PT and treating microbial infections, especially infection by  
 PT Staphylococcus aureus and Helicobacter pylori -  
 XX Claim 2; Page 2-3; 41pp: English.  
 XX  
 CC The present sequence is the gene encoding yphC polypeptide of  
 CC Staphylococcus aureus. The yphC coding sequence and protein are useful  
 CC for treating and diagnosing microbial infections such as infection caused  
 CC by S.aureus and Helicobacter pylori. In addition, the yphC coding  
 CC sequence and protein are useful for treating diseases such as  
 CC H.pylori-induced cancers, e.g. gastrointestinal carcinoma, gastric  
 CC ulcers, and gastritis. The present sequence was obtained from a library  
 CC of clones of chromosomal DNA of S.aureus in E.coli. The sequencing data  
 CC from two or more clones comprising overlapping S.aureus DNAs was used to  
 CC construct the present contiguous DNA sequence.  
 XX  
 SQ Sequence 1311 BP; 451 A; 184 C; 278 G; 398 T; 0 other;  
 Query Match 99.0%; Score 1292.2; DB 22; Length 1311;  
 Best Local Similarity 99.4%; Pred. No. 4.5e-287;  
 Matches 1297; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 ATGACTAAACCTATAGTACCTATTGTAGGTAGGCTTAATGTAGGTAAATCTACAATTTT 60  
 DB |||||  
 QY 1 ATGACTAAACCTATAGTACCTATTGTAGGTAGGCTTAATGTAGGTAAATCTACAATTTT 60  
 DB |||||  
 QY 61 AATAGAAATGTTGGAGAACGTTTCGATGTTGAAGATACGCCAGGTGTAACACAGAT 120  
 DB |||||  
 QY 61 AATAGAAATGTTGGAGAACGTTTCGATGTTGAAGATACGCCAGGTGTAACACAGAT 120  
 DB |||||  
 QY 121 CGTATTATTCTTCAGGTGAGTGGTTAACACATGATTTCAATATTATTGATACAGTGGT 180  
 DB |||||  
 QY 121 CGTATTATTCTTCAGGTGAGTGGTTAACACATGATTTCAATATTATTGATACAGTGGT 180  
 DB |||||  
 QY 181 ATTGAAATTTGGTATGACACCAATTCGAACACAAATTTAGCGGCGAGCAAAATCCGCCATA 240  
 DB |||||  
 QY 181 ATTGAAATTTGGTATGACACCAATTCGAACACAAATTTAGCGGCGAGCAAAATCCGCCATA 240  
 DB |||||  
 QY 241 GATGAAGCGGATGTTATTATTTTATGTTTAACGTGCGTGAAGGATTTGACACAAAGCCGAT 300  
 DB |||||  
 QY 241 GATGAAGCGGATGTTATTATTTTATGTTTAACGTGCGTGAAGGATTTGACACAAAGCCGAT 300  
 DB |||||  
 QY 301 GAAATGGTTCGCTCAAAATTTTATACAAATCTAAAAACCGGCTGCTATTACCGGTTAACAAA 360  
 DB |||||

DB |||||  
 QY 301 GAAATGGTTCGCTCAAAATTTTATACAAATCTAAAAACCGGCTGCTATTACCGGTTAACAAA 360  
 DB |||||  
 QY 361 GTAGATAATATGGAAATGCGGTACAGACGCTGATGATTTCTATTCAATAGGATTTGGTGAA 420  
 DB |||||  
 QY 361 GTAGATAATATGGAAATGCGGTACAGACGCTGATGATTTCTATTCAATAGGATTTGGTGAA 420  
 DB |||||  
 QY 421 CCGTATCCAATATACAGGCTCACATGGTTAGGTCTTGGTACTTCTTATAGATGCAATGTTT 480  
 DB |||||  
 QY 421 CCGTATCCAATATACAGGCTCACATGGTTAGGTCTTGGTACTTCTTATAGATGCAATGTTT 480  
 DB |||||  
 QY 481 TCTCATTTTGGTGAAGAGGAGGATCCTTATGATGAGATACAAATTCGACTATCCATT 540  
 DB |||||  
 QY 481 TCTCATTTTGGTGAAGAGGAGGATCCTTATGATGAGATACAAATTCGACTATCCATT 540  
 DB |||||  
 QY 541 ATTGGACGACAAACGCTAGTAAATCAAGTTTAGTAAATGCTATTATTAGTGAAGATCGC 600  
 DB |||||  
 QY 541 ATTGGACGACAAACGCTAGTAAATCAAGTTTAGTAAATGCTATTATTAGTGAAGATCGC 600  
 DB |||||  
 QY 601 GTTATCGTTTCTAATGTTGCGAGGACACGAGACGCTATTGATACAGATATAGTTAT 660  
 DB |||||  
 QY 601 GTTATCGTTTCTAATGTTGCGAGGACACGAGACGCTATTGATACAGATATAGTTAT 660  
 DB |||||  
 QY 661 GATGACAAAGATTTATGTTTAAATCGATCTGCTGATGCTGTAACGAAAGAAAGATATAT 720  
 DB |||||  
 QY 661 GATGACAAAGATTTATGTTTAAATCGATCTGCTGATGCTGTAACGAAAGAAAGATATAT 720  
 DB |||||  
 QY 721 GAATCAACTGAGAAATATTACGATTTAAGAGCTTTAAAGCGATTGAACGTTCAAAATGTT 780  
 DB |||||  
 QY 721 GAATCAACTGAGAAATATTACGATTTAAGAGCTTTAAAGCGATTGAACGTTCAAAATGTT 780  
 DB |||||  
 QY 781 GTTTTACTGCTTATGATGACAAAGGCTATTGTAACAGATTAACGTTGTCAGGA 840  
 DB |||||  
 QY 781 GTTTTACTGCTTATGATGACAAAGGCTATTGTAACAGATTAACGTTGTCAGGA 840  
 DB |||||  
 QY 841 TATGACATGAACAGGTTAAGCAGTCTGCTGATGCTGTAACGAAAGAAAGATGCTGAA 900  
 DB |||||  
 QY 841 TATGACATGAACAGGTTAAGCAGTCTGCTGATGCTGTAACGAAAGAAAGATGCTGAA 900  
 DB |||||  
 QY 901 AAAGATAGTAAACGATGAAGAAATTTGAAGATGAAGTACGTAAGAAATTTCAAAATTTTA 960  
 DB |||||  
 QY 901 AAAGATAGTAAACGATGAAGAAATTTGAAGATGAAGTACGTAAGAAATTTCAAAATTTTA 960  
 DB |||||  
 QY 961 GATTATGCCAAATTCCTTTGCTGCTGAAGAACCCACAGATTTACGTACATTTATTC 1020  
 DB |||||  
 QY 961 GATTATGCCAAATTCCTTTGCTGCTGAAGAACCCACAGATTTACGTACATTTATTC 1020  
 DB |||||  
 QY 1021 CCTTACATTAATGAACCAAGTGAACCAATTAACCAATTAACCAATTAACCAATTAACCAAT 1080  
 DB |||||  
 QY 1021 CCTTACATTAATGAACCAAGTGAACCAATTAACCAATTAACCAATTAACCAATTAACCAAT 1080  
 DB |||||  
 QY 1081 GAAAGTCTTACTGATCAATTTCCATGAACCTACACCAACAGAGTACAGCTTTG 1140  
 DB |||||  
 QY 1081 GAAAGTCTTACTGATCAATTTCCATGAACCTACACCAACAGAGTACAGCTTTG 1140  
 DB |||||  
 QY 1141 AATGCTCTTTATGCAACACAGTTCGTATAGAACCCGAGATTTGTTGTTATTTGTTAAT 1200  
 DB |||||  
 QY 1141 AATGCTCTTTATGCAACACAGTTCGTATAGAACCCGAGATTTGTTGTTATTTGTTAAT 1200  
 DB |||||  
 QY 1201 GATGTAGAATTAATGCAATTTTCTTATTAACGCTATTTAGAGAATCAAAATCCGCGCT 1260  
 DB |||||  
 QY 1201 GATGTAGAATTAATGCAATTTTCTTATTAACGCTATTTAGAGAATCAAAATCCGCGCT 1260  
 DB |||||  
 QY 1261 TTTGGTTTGAAGGTACACCAATTCATATTATAGCTGGAAGAGA 1305  
 DB |||||  
 QY 1261 TTTGGTTTGAAGGTACACCAATTCATATTATAGCTGGAAGAGA 1305  
 DB |||||

RESULT 4

AAS54997

ID AAS54997 standard; DNA; 1311 BP.

XX

AC AAS54997;

XX 13-FEB-2002 (first entry)  
 XX Staphylococcus aureus DNA for cellular proliferation protein #1309.  
 DE Antisense: ds; prokaryotic cellular proliferation gene;  
 XX antibiotic; antibacterial; drug design.  
 KW Staphylococcus aureus.  
 XX W0200170955-A2.  
 XX 27-SEP-2001.  
 PD 21-MAR-2001: 2001WO-US09180.  
 XX 21-MAR-2000: 2000US-191078P.  
 XX 23-MAY-2000: 2000US-206848P.  
 PR 26-MAY-2000: 2000US-207727P.  
 PR 23-OCT-2000: 2000US-242578P.  
 PR 27-NOV-2000: 2000US-253252P.  
 PR 22-DEC-2000: 2000US-257931P.  
 PR 16-FEB-2001: 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu RH;  
 XX WPI: 2001-611495/70.  
 DR P-PSDB: AAU37138.  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX Claim 27: Seq ID No 8634; Silpp: English.  
 XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence encodes an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 1311 BP: 452 A; 184 C; 278 G; 397 T; 0 other;

Query Match 98.9%; Score 1290.6; DB 23; Length 1311;  
 Best Local Similarity 99.3%; Pred No: 1;le-296;  
 Matches 1296; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ATGACTAAACCTATTAGTAGCTATTCTAGGTAGGCTTAATGTAGTAAATCTACAATTTT 60  
 DB 1 ATGACTAAACCTATTAGTAGCTATTCTAGGTAGGCTTAATGTAGTAAATCTACAATTTT 60

QY 61 AATAGTAAGTGTGGAGAACGCTGTTTCGATTGTTGAAGATACCCAGGCTGTAACACGAGAT 120  
 DB 61 AATAGTAAGTGTGGAGAACGCTGTTTCGATTGTTGAAGATACCCAGGCTGTAACACGAGAT 120

QY 121 CGTATTATTCTTCAGGTGAGTGTAAACACATGATTTCAATATTTATTGATACAGGTGGT 180  
 DB 121 CGTATTATTCTTCAGGTGAGTGTAAACACATGATTTCAATATTTATTGATACAGGTGGT 180

DB 121 CGTATTATTCTTCAGGTGAGTGTAAACACATGATTTCAATATTTATTGATACAGGTGGT 180  
 QY 181 ATTGAATATGGTGATGCACCATTTCCAAACACAAATTTAGAGCGCAGGAGAAATCGCCATA 240  
 DB 181 ATTGAATATGGTGATGCACCATTTCCAAACACAAATTTAGAGCGCAGGAGAAATCGCCATA 240  
 QY 241 GATGAAGCGGATGTTATTATTTTATGTTTAACTGCGTGAAGGATTGACACAAAGCGAT 300  
 DB 241 GATGAAGCGGATGTTATTATTTTATGTTTAACTGCGTGAAGGATTGACACAAAGCGAT 300  
 QY 301 GAAATGGTCCCTCAAAATTTTATACAAATCTAAAAACCCGGTCTGTTAGCCGTTAACAAA 360  
 DB 301 GAAATGGTCCCTCAAAATTTTATACAAATCTAAAAACCCGGTCTGTTAGCCGTTAACAAA 360  
 QY 361 GTAGATAATATGGAATGCGTACAGACGCTGTATGATTTCTATTCTATTAGGATTTGGTGA 420  
 DB 361 GTAGATAATATGGAATGCGTACAGACGCTGTATGATTTCTATTCTATTAGGATTTGGTGA 420  
 QY 421 CCGTATCCCAATATCAGGCTCACATGTTTAGCTCTTGGTCTGCTGACTTGTGTAGATGCAATTT 480  
 DB 421 CCGTATCCCAATATCAGGCTCACATGTTTAGCTCTTGGTCTGCTGACTTGTGTAGATGCAATTT 480  
 QY 481 TCTCATTTTGGTGAAGGAGGAAGATCCTTATGATGAAGATACAATTCGACTATCCATT 540  
 DB 481 TCTCATTTTGGTGAAGGAGGAAGATCCTTATGATGAAGATACAATTCGACTATCCATT 540  
 QY 541 ATTGGAGGACCAACGCTAGGTAAATCAAGTCTTATGATGAAGATACAATTCGACTATCCATT 600  
 DB 541 ATTGGAGGACCAACGCTAGGTAAATCAAGTCTTATGATGAAGATACAATTCGACTATCCATT 600  
 QY 601 GTTATGCTTTCTAATGTTTCAGGACACAGGAGACGCTATTTGATACAGAGTATAGTTAT 660  
 DB 601 GTTATGCTTTCTAATGTTTCAGGACACAGGAGACGCTATTTGATACAGAGTATAGTTAT 660  
 QY 661 GATGACACAGATTATGTTTTAATCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
 DB 661 GATGACACAGATTATGTTTTAATCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
 QY 721 GAATCAACTCAGAAATATTCAGTATTAAAGCGCTTTAAAGCGCTTTAAAGCGCTTTCAATGTT 780  
 DB 721 GAATCAACTCAGAAATATTCAGTATTAAAGCGCTTTAAAGCGCTTTAAAGCGCTTTCAATGTT 780  
 QY 781 GTTTTAGTGGTATTGATGACAGACAGGCTATTTGAAAGGATTAACAGGTTGTCAGGA 840  
 DB 781 GTTTTAGTGGTATTGATGACAGACAGGCTATTTGAAAGGATTAACAGGTTGTCAGGA 840  
 QY 841 TATGACATGAACAAGGTAAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 DB 841 TATGACATGAACAAGGTAAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 QY 901 AAAGATAGTAAACGATGAAGAAATTTGAAGATGAAGTACCTAAGAAATTCCTAATTTTAAAT 960  
 DB 901 AAAGATAGTAAACGATGAAGAAATTTGAAGATGAAGTACCTAAGAAATTCCTAATTTTAAAT 960  
 QY 961 GATTATGACAAATTCGCTTTTGTGCTGCTGAAGAACCCACAGATACCTGATACATTC 1020  
 DB 961 GATTATGACAAATTCGCTTTTGTGCTGCTGAAGAACCCACAGATACCTGATACATTC 1020  
 QY 1021 CCTTACATTAATCAAGCAAGTGAACCAACATAAAAACCGTGTTCAAAGTTCCTAATTTTAAAT 1080  
 DB 1021 CCTTACATTAATCAAGCAAGTGAACCAACATAAAAACCGTGTTCAAAGTTCCTAATTTTAAAT 1080  
 QY 1081 GAAGTCGTTTACTGATGCAATTTCCATGAACCCCTACACCAACAGACAAAGGTAGACGTTTG 1140  
 DB 1081 GAAGTCGTTTACTGATGCAATTTCCATGAACCCCTACACCAACAGACAAAGGTAGACGTTTG 1140  
 QY 1141 AATGTCCTTTTATGCAACACAAAGTTCGATATAGAACCCACCGACATTTGTTGTTGTTTAAAT 1200  
 DB 1141 AATGTCCTTTTATGCAACACAAAGTTCGATATAGAACCCACCGACATTTGTTGTTGTTTAAAT 1200  
 QY 1201 GATGTAGAATTAATGCAATTTTCTTATTAACCGCTATTATTAGAGAATCAATCCGCGCGCT 1260  
 DB 1201 GATGTAGAATTAATGCAATTTTCTTATTAACCGCTATTATTAGAGAATCAATCCGCGCGCT 1260



QY 1261 TTGCTTTGAAGCTACACCAATTCATATTATAGCTCGAAAGAGA 1305  
 Db 1261 TTGCTTTGAAGCTACACCAATTCATATTATAGCTCGAAAGAGA 1305

## RESULT 5

AAV74669  
 ID AAV74669 standard; DNA: 3621 BP.

AC AAV74669;

DT 16-MAR-1999 (first entry)

DE Staphylococcus aureus contig SEQ ID #358.

KW Computer readable medium; vaccine: S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.

OS Staphylococcus aureus.

PH Key Location/Qualifiers

FT misc\_feature 481..540

FT /\*tag- a

FT /\*note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence."

FT misc\_feature 2281..2340

FT /\*tag- b

FT /\*note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence."

PN EP786519-A2.

XX 30-JUL-1997.

XX 07-JAN-1997; 97EP-0100117.

XX 05-JAN-1996; 96US-0009861.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;

PI Rosen CA;

XX WPI; 1997-374922/35.

DR Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines

XX Claim 1: Page 1241-1243; 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences

CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.

XX Sequence 3621 BP; 1279 A; 442 C; 692 G; 1085 T; 123 other;

SO Query Match 94.3%; Score 1230.2; DB 18; Length 3621;  
 Best Local Similarity 94.6%; Pred. No. 1e-272;  
 Matches 1235; Conservative 1; Mismatches 69; Indels 0; Gaps 0;

QY 1 ATGACTAAACCTATAGTAGCTATTGTAGGTAGGCTAAATGTAGTAAATCTACAATTTT 60  
 Db 2130 ATGACTAAACCTATAGTAGCTATTGTAGGTAGGCTAAATGTAGTAAATCTACAATTTT 2189

QY 61 AATAGAAATAGTGGAGAACGCTGTTCCATGTTGTAAGATACGCCAGGTGTACACAGAT 120  
 Db 2190 AATAGAAATAGTGGAGAACGCTGTTCCATGTTGTAAGATACGCCAGGTGTACACAGAT 2249

QY 121 CGTATTATTCTTCAGGTGAGTGTAAACACATGATTTCAATATTATTATGATACAGGTGGT 180  
 Db 2250 CGTATTATTCTTCAGGTGAATGCTTAACACNNNNNNNNNNNNNNNNNNNNNNNN 2309

QY 181 ATTGAATTTGGTGATGCACCATTTCCAAACACAATTAGACGCCAGCAGCAAAATCCGCA 240  
 Db 2310 NNN 2369

QY 241 GATGAAGCGGATGTTATTATTTTATGTTAACTGCGTGAAGGATTGACACAAACGAT 300  
 Db 2370 GATGAAGCGGATGTTATTATTTTATGTTAACTGCGTGAAGGATTGACACAAACGAT 2429

QY 301 GAAATGGTGGCTCAAAATTTTATACAAATCTAAACAAACCGGTGCTATTACGCGTTAACAAA 360  
 Db 2430 GAAATGGTGGCTCAAAATTTTATACAAATCTAAACAAACCGGTGCTATTACGCGTTAACAAA 2489

QY 361 GTAGATAATATGGAATACCGGTACAGAGCTGTATGATTTCTATTTCATTAGATTTGGTGA 420  
 Db 2490 GTAGATAATATGGAATACCGGTACAGAGCTGTATGATTTCTATTTCATTAGATTTGGTGA 2549

QY 421 CCGTATCCAATATACAGGTGCACATGGTTTGGTGCTTTAGGTGCTTTAGATGACAGTTGTT 480  
 Db 2550 CCGTATCCAATATACAGGTGCACATGGTTTGGTGCTTTAGGTGCTTTAGATGACAGTTGTT 2609

QY 481 TCTCATTTTGGTGAAGAGAGATGCTTATGATGAAGATACATTCGACATATCCATT 540  
 Db 2610 TCTCATTTTGGTGAAGAGAGATGCTTATGATGAAGATACATTCGACATATCCATT 2669

QY 541 ATTGGACGACCAACCGTAGGTAAATCAAGTTTATGTAATGCTATTTTAGGTGAAGATCGC 600  
 Db 2670 ATTGGACGACCAACCGTAGGTAAATCAAGTTTATGTAATGCTATTTTAGGTGAAGATCGC 2729

QY 601 GTTATCGTTTCTTAATGTTGCAGGCAACAGAGAGCGCTATTGATACAGAGTATAGTTAT 660  
 Db 2730 GTTATCGTTTCTTAATGTTGCAGGCAACAGAGAGCGCTATTGATACAGAGTATAGTTAT 2789

QY 661 GATGACAAAGATTATGTTTTTAATCGTACTGCTGCTATGCGTAAAAAGGAAAAAGTATAT 720  
 Db 2790 GATGACAAAGATTATGTTTTTAATCGTACTGCTGCTATGCGTAAAAAGGAAAAAGTATAT 2849

QY 721 GAATCAACTGAGAAATATTCAGTATTAAAGAGCTTTTAAAGCGGATTGCAAGCTTCAATGTT 780  
 Db 2850 GAATCAACTGAGAAATATTCAGTATTAAAGAGCTTTTAAAGCGGATTGCAAGCTTCAATGTT 2909

QY 781 GTTTTAGTGGTTATTGATGCGAACAAGGCAATTTTGAACAAGATAAAGCTGTCGACGA 840  
 Db 2910 GTTTTAGTGGTTATTGATGCGAACAAGGCAATTTTGAACAAGATAAAGCTGTCGACGA 2969

QY 841 TATGCACATGACAAAGGTAAAGCAGTCTGATTGTCTGTAATAAATGGGATACTCTGAA 900  
 Db 2970 TATGCACATGACAAAGGTAAAGCAGTCTGATTGTCTGTAATAAATGGGATACTCTGAA 3029

QY 901 AAGATAGTAAACAGTGAACAAATTTGAAGATGAGTACGTAAAGAAATTTCAATTTT 960  
 Db 3030 AAGATAGTAAACAGTGAACAAATTTGAAGATGAGTACGTAAAGAAATTTCAATTTT 3089

QY 961 GATTATGCACAAATTCCTTTTGTCTGCTCTAAGAACCGCACAGGATTACGTACATTATTC 1020  
 DB |||||||  
 QY 3090 GATTATGCACAAATTCCTTTTGTCTGCTCTAAGAACCGCACAGGATTACGTACATTATTC 3149  
 DB |||||||  
 QY 1021 CCTTACATTAATGAAGCAAGTGAACACCAATAAAACCGTGTTCAAAGTTCAACTTTAAAT 1080  
 DB |||||||  
 QY 3150 CCTTACATTAATGAAGCAAGTGAACACCAATAAAACCGTGTTCAAAGTTCAACTTTAAAT 3209  
 DB |||||||  
 QY 1081 GAAGTCCTTACTGATGCAATTTCCATGAACCTCACCAACAGACAAAGGTAGACGTTG 1140  
 DB |||||||  
 QY 3210 GAAGTTGTTACTGATGCAATTTCCATGAACCTCACCAACAGACAAAGGTAGACGTTG 3269  
 DB |||||||  
 QY 1141 AATGTCCTTTTATGCCAACACAGTGTCTATAGAACCCGACATTTGTTGTTATTTAAAT 1200  
 DB |||||||  
 QY 3270 AATGTCCTTTTATGCCAACACAGTGTCTATAGAACCCGACATTTGTTGTTATTTAAAT 3329  
 DB |||||||  
 QY 1201 GATGTCGTTTAAATGCAATTTTCTTATAAACCGCTATTTAGAGAAATCAAAATCCGTCGCGCT 1260  
 DB |||||||  
 QY 3330 GATGTCGTTTAAATGCAATTTTCTTATAAACCGCTATTTAGAGAAATCAAAATCCGTCGCGCT 3389  
 DB |||||||  
 QY 1261 TTTGGTTTGAAGGTACACCAATTCATATATAGCTCGAAAGAGA 1305  
 DB |||||||  
 QY 3390 TTTGGTTTGAAGGTACACCAATTCATATATAGCTCGAAAGAGA 3434  
 DB |||||||

RESULT 6

ABN90883  
 ID ABN90883 standard; DNA; 1332 BP.  
 XX  
 AC ABN90883;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:346.  
 XX  
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy; gene; ds.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.  
 XX  
 PF 13-AUG-1998; 98US-0134001.  
 XX  
 PR 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PL Doucette-Stamm LA, Bush D;  
 XX  
 DR WPI; 2002-381255/41.  
 DR P-PSDB; ABP38338.  
 XX  
 PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 XX  
 PS Disclosure; SEQ ID 346; 267pp; English.  
 XX  
 CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35324 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.

XX Sequence 1332 BP; 465 A; 190 C; 267 G; 410 T; 0 other;  
 SQ  
 Query Match 70.6%; Score 921; DB 24; Length 1332;  
 Best Local Similarity 81.6%; Pred. No. 7.7e-202;  
 Matches 1065; Conservative 0; Mismatches 240; Indels 0; Gaps 0;  
 QY 1 ATGACTAAACCTATAGTAGCTATTTGTAGGTAGCGCTTAATGTAGGTAAATCTACAATTTT 60  
 DB |||||||  
 QY 22 ATGACTAAACCTATAGTAGCTATTTGTAGGTAAACCAATGTAGGTAAATCTACAATTTT 81  
 DB |||||||  
 QY 61 AATAGAATGATGAGAACCGTGTTCGATTTGTTGAAGATACCCAGGTGTACACAGAGAT 120  
 DB |||||||  
 QY 82 AATAGATGTTGCGCGAACCGTGTATCTATCTGTAAGAGATACGCCAGCGCTTACTCGTAT 141  
 QY 121 CGTATTTATCTTCAGGTGAGTGTGTTAAACACATGATTTCAATATATTATTGATACAGTGT 180  
 DB |||||||  
 QY 142 CGTATTTATCTTCGCGAGTGTGTTAACTCATGATTTTAACATCATTTGATACAGTGT 201  
 QY 181 ATTGAATGTTGATGACCATTTCCAAACACAAATTTAGAGCCGAGCGAANAATCCCATTA 240  
 DB |||||||  
 QY 202 ATTGAATGTTGATGATCTCTCTTTTCAAACTCAAACTCCGTGCACAGCGAANAATCAAT 261  
 QY 241 GATGAACGATGTTATTTATTTTATGTTAAGCGTGTGTAAGGATTTGACACAAAGCGAT 300  
 DB |||||||  
 QY 262 GAAGAAGCAGATGTCATCATTTTATGTCATGTTTATGAGAGAGGACTTTACACAAAGTCA 321  
 QY 301 GAAATGTTGCTCAAAATTTTATACAACTTAAACACCGTGTCTATAGCGGTTAAACAA 360  
 DB |||||||  
 QY 322 GAAATGTTGCTCAAAATGCTTTTAAATCTAAGAACCTGTTGATATAGCTGTGAATAA 381  
 QY 361 GTAGATAATGGAATGCTGTACAGAGTGTATGATTTCTTATTCATATTAGGATTTGGTGAA 420  
 DB |||||||  
 QY 382 GTTGATTAATCTGAAATGCTGTATATCTATGATTTCTTATTTAGGCTTTGGAGAT 441  
 QY 421 CCGTATCAATATCAGGTGTACATGTTTATGTTTGTGCTGTTTGTAGTGTGCTTGTAGT 480  
 DB |||||||  
 QY 442 CCATATCCTATTTCTGTTTCCATGATTTAGGATTTGGAGATTTGCTAGATGCAATGTT 501  
 QY 481 TCTCATTTGTTGAGAGAGAGAGATCTCTTATGATGAAGATACAAATTCGATATCCATT 540  
 DB |||||||  
 QY 502 GAAACCTTTAATAAGANTCAGAGAGATCTTTATGAGAAAGATACGATTCGTTTCCAT 561  
 QY 541 ATGGACGACCAACCGTAGTAAATCAAGTTTAGTAAATGCTATTTTAGGTGAAGATGCG 600  
 DB |||||||  
 QY 562 ATCGTAGACCTAATGTTGCTTAATCTAGCTTGTGCTCAATGCTATTTTAGGCGCAAGAC 621  
 QY 601 GTTATCGTTTCTAATGTTGAGGCAACAGAGAGCGCTTATGATACAGATATAGTTAT 660  
 DB |||||||  
 QY 622 GTTATGTTCTAATGTTGCTGTACAACTCGAGATGCCCTTGTATCCGAGTACTCTTAT 681  
 QY 661 GATGACAAGATTTATGTTTAAATCGATCTGTTGTTGCTGTTGTTGTTGTTGTTGTTGTT 720  
 DB |||||||  
 QY 682 GATGACAAGATTTATGTTTAAATGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 741  
 QY 721 GAATCAACTGAGAAATATTCAGTATTAAAGCGTTTAAAGCGATTGAACGTTTCAAAATGTT 780  
 DB |||||||  
 QY 742 GAATCGACTCAAAATATTCGTTTACGTCATTTAAAGCGATTGACGCTTCAGAGTA 801  
 QY 781 GTTTTACTGTTTATGATGACAGAACGATTTTGAACAGATTAATGATGTTGTTGTTGTTGTT 840  
 DB |||||||  
 QY 802 GTATTAGTATTGCTGATGCTGAACAGGTATTAATGAACAGATTAATGTTGTTGTTGTTGTT 861  
 QY 841 TATGCACATGAACAGGTAAAGCAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 900  
 DB |||||||  
 QY 862 TATGCACATGAGGAGGTAAAGCTATTTGTCATTTGTAGTAAATTAATTAATGGGATACAA 921  
 QY 901 AAACATAGTAAACAGTGAAGAAATTTGAAGATGAAGTACGTAAAGAAATTCCTCAATTTTA 960  
 DB |||||||  
 QY 922 AAACATAGTAAACAGTGAAGAAATTTGCTACTGATGATGTTAGAAATGAATTTCAATTTTA 981  
 QY 961 GATTATGCACAAATTCCTTTTGTGCTGTTAAGAACCGCACAGGATTACGTACATTATTC 1020  
 DB |||||||

Db 982 GATTATGCTCAATCGCGTTCGTATCAGCAAAAGAGGCGCTAAGATTAAAAACATTATTC 1041  
 QY 1021 CCTTACATTAATGAGCAAGTGAAGAACCAATAAAGACGTTTCAAGTTCACATTTAAAT 1080  
 Db 1042 CCTTATATCAATCAGCGATGGAATATATATAAGCGTGTCCAAAGTTCTACACTAAT 1101  
 QY 1081 GAATGCTGTACTGATGCAATTTCCATGAGACCCCTACACCAACAGAGAAAGGTAGACGTTG 1140  
 Db 1102 GAAGTGTGTACTGATGCAATTTCTATGAATCCAAACACCTACTGACAAAAGGTAGAGACTT 1161  
 QY 1141 AATGCTCTTTTATGCAACACAGTGTGCTATAGAACCCACCGACATTTGTTGTTATTTAAAT 1200  
 Db 1162 AATGCTATCTATACAACCTCAGGTGCAATTTGAATTTGAACCCACCGACATTTGTTGTTCAAT 1221  
 QY 1201 CATGTAGAAATTAATGCAATTTTCTTATAACCGCTATTAGCAATCAATCCGTCGCGCT 1260  
 Db 1222 GATGTTGAATTAATGCAATTTTCTTATAGGAGATATTAGAAAATCAATACGTAATGCT 1281  
 QY 1261 TTTGGTTTTGAAGTACACCAATTCATATTATAGTCGGAAGAGA 1305  
 Db 1282 TTTGGTTTTGAAGCAACACCTATTTCATATTATTATCCAAAGAAAAGA 1326

## RESULT 7

AAH54708/c

ID AAH54708 standard; DNA; 3269 BP.

XX AAH54708;

XX AC

XX DT 03-SEP-2001 (first entry)

XX XX

XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:4072.

XX Staphylococcus epidermidis SRI strain; infection; diagnosis;

XX Staphylococcus epidermidis; ds.

XX Staphylococcus epidermidis.

XX WO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000MO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX (GLAX ) GLAXO GROUP LTD.

XX Kimmerly WJ;

XX WPI; 2001-316495/33.

XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis;

XX useful for vaccinating against infections, e.g. endocarditis -

XX Claim 8; Page 1757-1759; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides

XX (I), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.

XX (I) and (II) can have antibacterial activity and therefore can be used

XX in vaccination. The nucleic acids (I) may be used to produce the

XX S. epidermidis polypeptides (II) via the production of vectors

XX containing them which are used to produce hosts cells which express the

XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

XX used to vaccinate subjects and to raise antibodies against the bacteria.

XX The polypeptides may also be used to assay for other inhibitors of their

XX activity and therefore identify compounds that may be used for the

XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to

XX AAH55090 represent specifically claimed S. epidermidis genomic DNA

XX polynucleotide sequences from the present invention. AAH55091 to

XX AAH55098 represent oligonucleotide sequences and primers which are used

XX in the exemplification of the present invention.

XX N.B. The present invention specifically claims all the polynucleotide

CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 3269 BP; 986 A; 616 C; 451 G; 1216 T; 0 other;

Query Match 70.5%; Score 919.4; DB 22; Length 3269;  
 Best Local Similarity 81.5%; Pred. No. 2.3e-201;  
 Matches 1064; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 1 ATGACTAAACCTATATAGTAGCTATTGTAGGTAGCGCTTAATGTAGTAAATCTACAAATTTT 60

Db 2804 ATGACTAAACCGATATAGCAATTTGTAGGAAAACCAATGTAGGAAAATCTACAAATTTT 2745

QY 61 ANTAGAATTTGGAGAACGTTTTCGATTGTGAAGATACGCCAGGTGTACACAGAT 120

Db 2744 AATAGAGTTGTGGCGAACGTTATCTATCTAGAGATACGCCAGGCTTACTCGTGT 2685

QY 121 CGTATTATCTTCAGGTGAGTGGTTAACACATGTTTCAATATTATTATGATACAGGTGT 180

Db 2684 CGTATTATCTTCAGGTGAGTGGTTAACATGTTTCAATGTTTAACTCATGATCATTTGATACAGGTGT 2625

QY 181 ATTGAATTTGGTGTGATGCCACATTTCCAAACAAATTTAGAGCGCAGAGAAATCGCCATA 240

Db 2624 ATTGAATTTGGAGATGCTCTTTTCAAACTCAATTCGTCGACAGGCAAGATACCAAT 2565

QY 241 GATGAAGCGGATGTTATTTATGTTAACTGCGTGAAGGATTTGACACAAAGCGAT 300

Db 2564 GAAGAAGCAGATGTCATCATTTTATGCTCAATGTTAGAGAGAGGACTTACACAAAAGTGAC 2505

QY 301 GAAATGGTGGTCAAAATTTTATACAAATCTAAAAAACCGTGTATTTAGCGGTAAACAAA 360

Db 2504 GAAATGGTGGCACAATGCTTTTATAAATCTAAGAACCTGTTGTTATAGCTGTGAATAA 2445

QY 361 GTAGATAATATGGAATTCGTCAGACGCTGTATGATTTCATATCTATTAGGATTGGTGAA 420

Db 2444 GTTGATAATCTGGAATTCGTCAGACGCTGTATGATTTCATATCTATTAGGATTGGTGAA 2385

QY 421 CGTATCCAAATATCAGGTCACATGTTTATGTTAGTCTTGGTGACTTTTATAGATCAGTGT 480

Db 2384 CCATATCTATTTCTGGTTCACATGTTAGGACTTGGAGATTTGCTAGATCAGTGTGT 2325

QY 481 TCTCATTTTGGTGAAGAGAGATCCTTTATGTAAGATACAAATTCGACTATCCAT 540

Db 2324 GAAAACTTTAATAAGATCAGAAGATCCTTTATGTAAGATACAAATTCGACTATCCAT 2265

QY 541 ATTGACGACCAACAGGTAGGTAAATCAAGTTTATGTAAGTCTATTTTAGGTGAAGATCGC 600

Db 2264 ATCGGTAGACCTAATTTGGTAAATCTAGCTTGGTCAATTTAGGCGGAAGAGCT 2205

QY 601 GTTATCGTTTCTAATGTTGAGGACCAACGAGAGAGCTATTGATACAGAGTATAGTTAT 660

Db 2204 GTTATGTTCTAATGTTGAGGACCAACGAGAGAGCTATTGATACAGAGTATAGTTAT 2145

QY 661 GATGACAAGATTTATGTTTAAATCGATCTGCTGTTATGCTGTTAAAAAGGAAAGATATAT 720

Db 2144 GATGACAAGATTTATGTTTAAATCGATCTGCTGTTATGCTGTTAAAAAGGAAAGATATAT 2085

QY 721 GAATCAACTGAGAAATATTCAGTATTAAAGAGCTTTAAAAAGCGATTGAAACGTTCAATGTT 780

Db 2084 GAATCGACTGAAAAATATTCCTGTTATACGTGCTATTAAGAGCGATTGAGCGTTTCAGAGTA 2025

QY 781 GTTATTAGTGGTATTGTCAGAACAAAGCGATTATTGAAACAGATAAACGTTTTCACGA 840

Db 2024 GTTATTAGTGGTATTGTCAGAACAAAGCGATTATTGAAACAGATAAACGTTTTCACGA 1965

QY 841 TATGCACATGAACAGGTAAAGCAGTCTGATTGTCGTAATTAATTAATGGGATACTGTGAA 900

Db 1964 TATGCACATGAACAGGTAAAGCAGTCTGATTGTCGTAATTAATTAATGGGATACTGTGAA 1905

QY 901 AAGATATCAAAACGATGAAGAAATTTGAAGATGAAGTACGTAAAGAAATTTCAATTTTAA 960

Db 901 AAGATATCAAAACGATGAAGAAATTTGAAGATGAAGTACGTAAAGAAATTTCAATTTTAA 960











1132	ACAGCTTTGANTGCTTTTATGCAACACAGAGTTGCTATAGAACACCGACATTTGTGTGA	1191
QY		
781	AGAGACCTTAATGATTCTATACACTCAGGTTGCAATTGAACACCGACACATTTGTAGTA	840
DB		
1192	TTTGTTAATGATGATAGAAATTAATGCAATTTTTCTTTATAAACGCTATTTAGAGAAATCAAATC	1251
QY		
841	TTTGTCAATGATGTTGAAATTAATGCAATTTTTCTTTATAGGAGATATTTAGAATAATCAAATA	900
DB		
1252	CGTCCCGCTTTTGGTTTTGAAGGTACACCAACTTATATATATAGCTCGAAAGAGA	1305
QY		
901	CGTAATGCTTTTGGTTTTGAAGGAGACACCTTATCATATATTTCCAAGAAAAAGA	954
DB		

**RESULT 12**

ABN68457  
ID ABN68457 standard; DNA; 1308 BP.

AC ABN68457;

DT 01-JUL-2002 (first entry)

DE Streptococcus polynucleotide SEQ ID NO 4827.

Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae; KW  
group A streptococcus; Streptococcus pyogenes; antibacterial; gene; KW  
antiinflammatory; infection; vaccine; meningitis; gene therapy; ds. KW

OS Streptococcus agalactiae.

XX  
PN  
WO200234771-A2.

02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000: 2000GB-0026333.

PR 07-MAR-2001; 2001GB-0005640.

PA (CHIR-) CHIRON SPA.

XX  
XX  
(CONT) INST GENOMIT

PI Tettelin H;

DR WPI: 2002-352536/38.

XX  
XX  
P-130B; ABP:

PT New Streptococcus pr  
PT or disease caused by

XX  
Pl for detecting a c

PS  
XX  
Claim 7; Page 3644-3645; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus/CBS (Streptococcus agalactiae) or group A streptococcus/

CC (Streptococcus pyogenes), comprising one of 5483 sequences

CC activity. (I), nucleic acids encoding (I)

the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*. Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins.

Sequence 1308 BP; 389 A; 204 C; 290 G; 425 T; 0 other:

Query Match 46.0%; Score 600; DB 24; Length 1308;  
Best Local Similarity 67.5%; Pred. No. 3.9e-128;  
Matches 875; Conservative 0; Mismatches 415; Indels 6; Gaps 2;

Qy	9	ACCTATAGTACGTATTTGTAGTAGGCGCTAAATGTAGGTAAATCTACAAATTTTTTAAATAGAAAT	68
Db	9	ACCTACAGTTGCTATTTGTTTGGTCTGCGGAATGTTGGAAAATCAAACTATATTTAAACAGAA	68
Qy	69	AGTTGGAGAACGTGTTTTCGATTTGTTGAAGATACGCCAGGTGTAAACACGAGATCGGTATTTTA	128
Db	69	TGCAGGAGAACGTATTTCCATTTGTTGAGGACGTTGAAGGAGTCACTCGTGTATCGTATCTA	128
Qy	129	TTCTTCACGTGAGTGGTTTAAACATGATTTTCAATATTTATTGTATACAGGTGGTATTGTGAAT	188
Db	129	TACACGGGTGAGTGGCTAAATCGTAAAGTTTCTCATTTAATTGATCTGTGTGGGATTGATGA	188
Qy	189	TGGTGATGACACATTTCCAAACACAAATTTAGAGCGCAGGAGAAATCGCCATAGATGGAAC	248
Db	189	TGTTGATGACACGTTTATGGAGCAAAATTAAGCACCAAGCAGACATTTGCTATGACAGAGGC	248
Qy	249	GGATGTTATTTATTTATGTTTAACTGTCGCTGAAGGATTTGACACAAAGCGATGAAATGGT	308
Db	249	TGATCTTATTTGTTGTTGTCGGGTAAAGAAAGGTGTGACTGTGCAGATGAGTATGT	308
Qy	309	CGCTCAAAATTTATACAAATCTAAAAAACCGTCGTATTTAGCGGTTTACAAAGTAGATAA	368
Db	309	CTCAGCTATCTTTATAGACCAATAAACCCGTTATTTTACGTCGTTAATAAGGTAGATAA	368
Qy	369	TATGAAATGGCTACAGACGCTGTATGATTTCTATTCAATTAGGATTTGGTGAACGCTATCC	428
Db	369	TCTTGAGATGCGCAATGATATTACGATTTTCTACTCTCTGTTTGAAGGATCCTTATCC	428
Qy	429	AAATATCAGGTCACATGGTTTATAGTCTTGTGTGACTTTTGTAGATGCAAGTGGTTCTCATTT	488
Db	429	CTTGCTATCAGTTTCATGGTATTTGGTACTGGTGATATATTGGATGCTATTTTGGAAATCT	488
Qy	489	TGGTGAAGAGAGAGATCCTTTATGATGAAGATACAAATTCGACTATCCATTTATTTGGACG	548
Db	489	TCCGTTGAAGAGAGAAAA---TGAAAAATCCAGATATTTATTCGTTTTTAGTTGATTGGACG	545
Qy	549	ACCAAACGTAGGTAAATCAAGTTTTAGTAAATGCTATTTTAGTGAAGATCCCGTTATCGT	608
Db	546	TCCAAATGTTGGGAAATCTAGTTTAATTAATGCCATCCTTGGGGAAGACCGTGTATTGTC	605
Qy	609	TTCTTAATGTCGAGGACACGAGAGCGCTATTGTATACAGATAGTACTTAAT---GATGG	665
Db	606	TAGCCCAAGTTGCTGTAATGCGCGATGCTATTGATCTAAATTCGTTGATACCCAGGG	665
Qy	666	ACAAGATATGTTTTTAATCGACTACTGCTGTATGCGTAAAAAAGGAAAAGTATATGAATC	725
Db	666	CGAAGAGTATACTATGATTTGATACTGCTGCTGATTCGCTAAATCTGGTAAAGGTATATGAAA	725
Qy	726	AACTCAGAAATATTACGATTTAAGAGCTTTTAAAAGCGATTGAAACGTTCAAAATGTTGTTTT	785
Db	726	CACAGAGAAATATTCTGTTATGCTTCTATGCGCGCCATCGACCCGCTCAGATGTTGTTCT	785
Qy	786	AGTGGTTATTGATGCAGAACAGGCCATTATTGAAACAAGATAAACCGTTTGCAGGATATGC	845
Db	786	TATGCTGATTAATGCAGAGGAAGTATTCGTGAGTATGACAAGCGTATTCGACGGGTTTGC	845
Qy	846	ACATGAACAAAGGTAAAGCAGTCGTTGTTGCTAAATTAATGGGATCTGTGCGAAAAGA	905
Db	846	ACATGAAACTGCTAAAGCTATTATTATTTGTAGTTAATAATGGGATCTATCGAAAAGA	905
Qy	906	TAGTAAACGATGAAGAAATTTGGAAGATGAAGTACCTTAAGAAATTTTCCAAATTTTAGATTA	965
Db	906	CAATCACTACTGCTAGTCAGTGGGAAGCTGATATCTGTGATAACTTCCAAATTTCTGCTTA	965
Qy	966	TGCACAAATGCTTTTGTGCTGCTAAAGAACCCACAAAGATTACGTACATTTATTTCCCTTA	1025
Db	966	TGCGCAATATCTTTGTTCTGCACAGACAAAGCAACGTTTGGCAAGATTTACCAGATAT	1025





DB 1589796 CAATCATCTGTCAGTCAGTGGGAAGCATATATAGTGTAACTTCCAAATCTTGTCTTA 1589797  
 QY 966 TGCACAAATTCGTTTGTGCTGCTAAAGACGCAAGATTACGTATATTCCTTA 1025  
 DB 1589736 TGCACCAATTCATCTTGTGCTGCTGAGACAAAGCAAGCTTGGCACAAGTTACAGATAT 1589677  
 QY 1026 CATTAAATGAAGCAAGTGAACCATTAACCAAGCTGTTCAAAGTTCAACCTTTAAATGAAT 1085  
 DB 1589676 CATTAAACGTATTAGCGCAAGTCAAAACCAAGCTATTCGCTGCTGTATTAATGATGT 1589617  
 QY 1086 CGTTACTGATGCAATTCCTGATGACCTACACACACCAAGGTAGACCTTTGAATGT 1145  
 DB 1589616 GATTATGATGCTATGCTATTAATCCACACCACTGTAAAGCAAGCTTTAAAGT 1589557  
 QY 1146 CTTTATGACCAACAAAGTGTGCTATAGACCAAGCTATTTGTTGTTTAAATGATGT 1205  
 DB 1589556 TTTCTACGCAACTCAAGTTGCTGTAAACCAACCACTTTGCTGTTGTTTAAACGAAGA 1589497  
 QY 1206 AGAATTAATGCAATTTCTTATTAACGCTATTTAGAGAAATCAATCCGTCGCTTTGG 1265  
 DB 1589496 AGAATGATGCACTTTTCATATCTTCTTCTTAAATCAATCCGTCGAGCAATTTGT 1589437  
 QY 1266 TTTTGAAGCAACCAATTCATATATATAGCTCGAAA 1301  
 DB 1589436 TTTTGAAGCAACCAATTAATCTAATAGCAGTAA 1589401

RESULT 14  
 ABK75008  
 ID ABK75008 standard; DNA: 1311 BP.  
 XX  
 AC ABK75008;  
 XX  
 DT 13-AUG-2002 (first entry)  
 XX  
 DE Bacillus licheniformis genomic sequence tag (GST) #2299.

XX Differential gene expression; genomic sequenced tag: GST;  
 KW altered culture condition; environmental stress;  
 KW physiological provocation; ds.

XX Bacillus licheniformis.

XX WO200229113-A2.

XX 11-APR-2002.

XX 05-OCT-2001; 2001WO-US31437.

XX 06-OCT-2000; 2000US-0680598.

XX 27-MAR-2001; 2001US-279526P.

XX (NOVO) NOVOZYMES BIOTECH INC.

XX (NOVO) NOVOZYMES AS.

XX Berka R, Clausen IG;

XX WPI; 2002-416684/44.

XX Monitoring differential expression of several genes in first Bacillus  
 PT cell relative to expression of same genes in one or more second  
 PT Bacillus cells, by using substrate containing Bacillus genomic  
 PT sequenced tag array

XX Claim 4: SEQ ID NO 2299; 200pp; English.

XX The invention describes a method of monitoring differential expression of  
 CC genes in a first Bacillus cell relative to expression of the genes in  
 CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
 CC isolated from Bacillus cells to a substrate containing array of Bacillus  
 CC genomic sequenced tags (GST), examining the array, and determining  
 CC relative gene expression by an observed hybridisation reporter signal of

CC a spot in the array. The method is useful for measuring the expression of  
 CC genes in a first Bacillus cell relative to expression of the same genes  
 CC in one or more second Bacillus cells. The method is useful for monitoring  
 CC global expression of several genes from a Bacillus cell, discovering new  
 CC genes, identifying possible functions of unknown open reading frames and  
 CC monitoring gene copy number variation and stability. Monitoring changes  
 CC in expression of genes may be used to provide a representation of the way  
 CC in which Bacillus cells adapt to changes in culture conditions.  
 CC environmental stress or other physiological provocation. Extensive  
 CC follow-up characterisation is unnecessary, when one spot on an array  
 CC equals one gene or one open reading frame, since sequence information is  
 CC available. This sequence represents a genomic sequence tag (GST) used in  
 CC the method of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1311 BP; 376 A; 279 C; 353 G; 303 T; 0 other;

Query Match 45.98; Score 599.4; DB 24; Length 1311;  
 Best Local Similarity 66.28; Pred. No. 5.3e-128;  
 Matches 864; Conservative 0; Mismatches 441; Indels 0; Gaps 0;

QY 1 ATGACTAAACCTATAGTAGCTATTAGTAGTAGCCCTAATAGTAGTAAATCTACAATTTT 60  
 DB 1 ATGGGAAAACCTGCTGCTAGCCATAGTTGGAAGACCCCAATGTGGCAAGTCCACGATCTTT 60  
 QY 61 AATAGAATAGTTGGAGAACGCTGTTTCGATTGTTGAAGATACGCCAGGTGTAAACAGAGAT 120  
 DB 61 AACAGGATCGCGGCGGAAGAAATTTTCGATTGTGGAAGATACACCTGGCGACGCGGAT 120  
 QY 121 CGTATTATCTTCAGGTGAGTGGTTAAACACATGATTTCATATTTATGATACAGGTGGT 180  
 DB 121 CGGATATACAGCTCTGCGGAGTGGCTGAATACAGACTTCACAGCTGATGATACAGCGGA 180  
 QY 181 ATTGAATTTGGTATGACCCATTTCCAAACACAAATTAGAGCCGACGAGAAATGCCATA 240  
 DB 181 ATCGAAGTCGCGGATGAGCGTTTTCGCGCAGATTCGCGCATCAGCCGAGATGCCATG 240  
 QY 241 GATGAAGCGGATGTTATTTATTTATGTTTACGTCGTTGAAGGATTCGACCAAGCGAT 300  
 DB 241 GAAGACGAGATGATTATTTTCATGACGACGCGCGGAGGCGTCCAGCCGCTGAT 300  
 QY 301 GAAATGTCGCTCAAAATTTTATACAAATCTAAATAACCGGTGCTATTAGCGGTTAAACAA 360  
 DB 301 GAAGAAGTCGCGGAAAATTTTATACCGGTACAAAACCAACCCGCTGCTCGCGGTTAAATA 360  
 QY 361 GTAGATAATATGGAATTCGCTACAGACGCTGTATGTTCTATTCTATTAGGATTTGGTGAA 420  
 DB 361 GTGGATAATCTGGAATGAGACCAACATTTACGATTTTATGCCCTCGGCTTTGGCGAG 420  
 QY 421 CCGTATCCAAATCAGGGTCACATGGTTTAGGTCTTTGGTGACTTGTAGATGCAATGTTGT 480  
 DB 421 CCGTTCCCGATTTTCATGGACGCTGCTTGGTTCGGCGATCTGCTTTGACGCGGTGAGC 480  
 QY 481 TCTCATTTTGTGAAGAGGAAGATCTCTTATGATGAAGATACAAATTCGACTATCCATT 540  
 DB 481 GAACATTTTAAACATATACCCGAGACGAGATGAGGATGAGTCTGTTCAATTTTGCCTC 540  
 QY 541 ATTGGACGACCAACAGCTAGGTAATCAAGTTTAGTAAATGCTATTTTAGGTGTAAGATCGC 600  
 DB 541 ATCGGCGCGGCAACGTCGGAAAATCTCTCCCTTGTAAACGCGCATGATTTGGAGAAAGCGC 600  
 QY 601 GTTATCGTTTCTAATGTTGCGAGGACACGACGAGACGCTATTGATACAGATATAGTTAT 660  
 DB 601 GTTATCGTCAGCAATATTGCGGGAACGACGCGACGCCATTTGACACAGGTTTACATAC 660  
 QY 661 GATGACCAAGATGTTTAAATCGATCTGCTGATCCGCTATCCGTAATAAAGGAAAGATATAT 720  
 DB 661 AATCAAGACACTTTGTCATCGTGGATACAGCGCGGATGAGAAAAGGAAAGTATAT 720  
 QY 721 GAATCAACTGAGAAATATTTCAGTATTAAAGAGCTTTAAAGCGATTGAAAGCTTCAAAATGTT 780

Db 721 GAGGCACTGAAATAACAGCGTCTCCGAGCTTTAAAGCGGATCGACCGTTTCAGAGTC 780  
 Qy 781 GTTTTAGTGGTATTGATGACAGCAAGGCATATTGACAAAGATAAAGCTGTTCAGGA 840  
 Db 781 GTCTCGGCTGTGGACGGCGAAGAGGCAATTATTGAGCAGGACAGCGGATTCGCGGC 840  
 Qy 841 TATGCACATGAACAGGTAAGCAGTCTGATTTGCTGTAATAAATGGGATCTGTGGA 900  
 Db 841 TATGCCATGAAGCAGGAAAGCTGTCTCATCGTCTGTAACAAATGGGATCGAGTCAA 900  
 Qy 901 AAAGATAGTAAACAGCATGAAGAAATTTGAAGATGAAGTACGTAAAGAAATTCCAATTTTA 960  
 Db 901 AAAGACGAGCGGACGATGAAGAAATTTGAAGAAAGGTTAGAGACCACTTTCATTTTGG 960  
 Qy 961 GATTATGACAAATTTGCTTTGCTGCTTAAGAACGCAAGATTTAGCTACATATTTC 1020  
 Db 961 GACTACGCGGCTGTGTTGTTTATGTCGCTTGGCTTAAGAAAGCGCATCCATACGCTGATG 1020  
 Qy 1021 CCTTACATTAATGAAGCAAGTGAACCAATTAAGAAAGCGTGTTCAAAGTTCAACTTTAAAT 1080  
 Db 1021 CCTCGGTCATGACGCGGAGCAACCACTTCGATGAGGCTGCAGACCAATATCTTAAAT 1080  
 Qy 1081 GAAGTCGTTACTGATGCAATTTCCATGACCCCTACACCAAGCAAGAGTAGAGTTTG 1140  
 Db 1081 GATATCATATTGATGCGGCTGCTATGACCCGCGCGCGCTAATGGAACCGGTTA 1140  
 Qy 1141 AATGCTTTTATGCAACAAAGTTGCTATAGAACCCGACATTTGCTGATTTGTTAAAT 1200  
 Db 1141 AAAATTTATATGCGACAAAGTGGCTGTGAAGCCGCGGACATTTGCTGATTTGTCAT 1200  
 Qy 1201 GATGATGAATTAATGCATTTTCTTATTAACGCTATTAGAGAAATCAAAATCGGTCGCT 1260  
 Db 1201 GACCCGTGATGATGATTTTCTTACGAAAGCTTTTGGAAACCGGATACGCGAGCGC 1260  
 Qy 1261 TTTGCTTTGAGGTACACCAATTCATATTAGTCCGAAGAGA 1305  
 Db 1261 TTCGATTTGAAGGAACCAACATCAGGATTTTCGACAGAGCAAGA 1305

## RESULT 15

ID ABA90521  
 XX ABA90521 standard; DNA; 2365589 BP.

AC ABA90521:

XX 16-MAY-2002 (first entry)

XX Genomic sequence of *Lactococcus lactis* IL1403.

DE Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.

KW *Lactococcus lactis* IL1403.

OS FR2807446-A1.

XX 12-OCT-2001.

XX 11-APR-2000; 2000PR-0004630.

XX 11-APR-2000; 2000PR-0004630.

XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.

PA Bolotline A, Sorokline A, Renault P, Ehrlich SD;

PI wPI; 2002-043418/06.

XX New nucleotide sequence useful in the identification of *Lactococcus*  
 PT *lactis* and related species -

XX Claim 1: SEQ ID 1: 2504pp; French.

CC The present invention is related to a *Lactococcus lactis* nucleotide  
 CC sequence (ABA90521) and related proteins (ABBS3300-ABBS5621). The  
 CC nucleic acid sequence is useful in the detection and/or amplification of  
 CC related species. The proteins of the invention are useful for the  
 CC biosynthesis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic bacteria, particularly useful in the  
 CC production of yogurt and cheese.  
 CC Note: The sequence data for this patent is based on equivalent patent  
 CC W0200177334 (published 18-OCT-2001) which is available in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;

SQ Query Match 44.6%; Score 582.4; DB 24; Length 2365589;

Best Local Similarity 66.5%; Pred. No. 3e-123;

Matches 867; Conservative 0; Mismatches 431; Indels 6; Gaps 2;

Qy 1 ATGACTAAACCTATAGTAGCTATTGTAGTAGGCTAATAGTAGTAATCTACAATTTT 60  
 Db 761403 ATGAGCTTACCTACAGTAGCCATCGTTGGCGTCCAAATGCGGAAATCAACGATATTC 761462  
 Qy 61 AATAGAAATAGTTGGAGAACGTTTTCGATTCTCAAGATACGCCAGCTGAACACGAGAT 120  
 Db 761463 AACCGTATTGACGAGAACGCAATCTCAATGTCGAGACATTCGAGGTGTAACGCTGAC 761522  
 Qy 121 CGTATTATTCTTCAGGTGAGTGGTAAACACATGATTTCAATATTATTGATACAGGTGT 180  
 Db 761523 CGTATCTATGCCACAGGAGAAATGTTAAACCGTAAATTTCAATATCATTCGATACAGGTGG 761582  
 Qy 181 ATTGAATTTGGTGTATGTCACCATTTCCAAACACAAATTAGAGCGGAGGACAGAAATCGCCATA 240  
 Db 761583 ATTGAATTTTACAGCAACCTTTTCATGATCTGAAATTCGAGCGCAAGCTGAAATTCGATG 761642  
 Qy 241 GATGAGCGGATCTTATTATTATTTTATGTTAACTGCTGCGTGAAGGATTCACAAACCGAT 300  
 Db 761643 ACCGAACGAGATGTTATCATCGCTGTTGTCGCGGAGAACAGGATTCACAGATCCGAT 761702  
 Qy 301 GAAATGGTCGCTCAAAATTTTATACAAATCTAAAAACCGGTCGTATTAGCGGTTAACAAA 360  
 Db 761703 GAAGCGGTTGCTTAATATTCTTATCGTACAGATTAACCCAGTATTCTGTTGTCATATAA 761762  
 Qy 361 GTAGATAATATGAAATCGCTACAGACGTGTATGATTCTTATTCTATTAGGATTTGTTGAA 420  
 Db 761763 GTTGATAACCCAGAACGTCGAATGGAATCTTCTGATTTCTACTCACTTGGACGT 761822  
 Qy 421 CGTATCCAAATATCAGGTCACATGGTTAGGCTCTTGGTCACTTGTAGATCAGTTGTT 480  
 Db 761823 CTTATCTGTTTACGCGGTTTCAGCGGATTCGCGGATTCGCGGATTCGCTTGAATCAATCGTT 761882  
 Qy 481 TCTCATTTTGTGAAGAGAAAGATCTCTTATGATGAAGATACAATTCGACTATCCATT 540  
 Db 761883 CAATAATCTTCTTAATGAATTTGAAGA---AGAAAATGAGAAATGTCATCAAGTTCAGCTTG 761939  
 Qy 541 ATTGGACGACCAACAGCTAGGTAAATCAAGTTTGTAGTAATGCTATTATTAGTGGAAGTCG 600  
 Db 761940 ATTGGCGCGCAACAGCTTGGGAAATCATCTCTTAAATGCTATTCTTGGCGAAGACCGC 761999  
 Qy 601 GTTATCGTTTCTTAATGTTTGGAGGACACAGACACCGCTATTGATACAGATATAG---T 657  
 Db 762000 GTAAATGCAAGTCCAAATTCGAGGACCAACTCGTATGCGGATTCGACACACACATTTGTCGAT 762059  
 Qy 658 TATGATGGAACAAGATTATGTTTTTAATCGATACCTGCTGTTATGCGTAAAAAGGAAAGTA 717  
 Db 762060 TCTGAAGACCAAGAATTTGTTATGATTGATACCCCGGAATGCGTAATTCAGGAAAAAT 762119  
 Qy 718 TATGAATCACTGACAAATATTAGATTTAAGAGCTTTTAAACGCTTTCAGCTTCAAT 777  
 Db 762120 TACGAAATATCTGAAAAATATCTAGTCATGCGTCCCATGCGTCCATTCGCTTCAGAT 762179  
 Qy 778 GTTGTGTTTAGTGGTTATTGATGACGAAAGGCAATTTATGCAACAAGATAAACGTGTGCA 837  
 Db 762180 ATCGTTTTGATGTTGATTAACGCTTGAGGAAGGAATTCGCGAATATGATGCGTATCGCT 762239

OY 838 GGATATGCACATGAACAGGTAAAGCAGTCGTCGATTCGTAAATAAATGGGATACCTGTG 897  
Db 762240 GGATTTGCTCACCAGACTGGTAAGGAATTTTGATGTGTGTCATTAATGGGATACCCTT 762299  
OY 898 GAAAGATAGTAAACAGATGAAGAAATTTGAAGATGAAGTACGTAAGAAATTCCTCAATTT 957  
Db 762300 GAGAAAGATAACGATACAAATGAAGAACTTTGAATTAGAAATTCGGACAAAGTTCAAATTC 762359  
OY 958 TTAGATTATGCACAAATTCCTTTGTCGTCGTAAGAACGACAAAGATTACGTACATTA 1017  
Db 762360 CTTGATATGCACCAATTCCTTTATGTCGTAAGACAGACACGTCGTAATTAATC 762419  
OY 1018 TTCCCTTACATTAATGAAGCAAGTGAAGAACCAAAAAACGTCGTTCAAAAGTTCACATTTA 1077  
Db 762420 CCAGATATGATTAAGAAATTCATCATGCACAAAACCTTACGTAATTTCAAGTTCGTGTTTG 762479  
OY 1078 AATGAAGTCGTTACTGATGCAATTTCCATGAACCCCTACACCAACAGACAAAGGTAGACGT 1137  
Db 762480 AATGATGTCATTTATGATGCTGTAGCGATTATCCAACTCCGACGGACAAAGGAAACGT 762539  
OY 1138 TTCAATGCTCTTTTATGCAACACAAAGTTGCTATAGAACCCAGACATTTGTTGATTTGTT 1197  
Db 762540 TTGAAGATTTTCTATGCAACACAAAGTAGCGATTAAACCCAACTTTTGTGTTTGTGTC 762599  
OY 1198 AATGATGTAGAATTAATGCAATTTTCTTATAAACGCTATTTAGAGAATCAAAATCCGTGCC 1257  
Db 762600 AATGAAGAAGAACTCATGCACCTCTCATATCTTCGGTTCTTGAATAATCAAAATCCGTAAA 762659  
OY 1258 GCTTTTGGTTTTGAAGGTACACCAATTCATATTATAGCTCGAAA 1301  
Db 762660 GCCTTGTGTTTTGAGGACACCAAGTTTCATTTGATTCACGTA 762703

Search completed: September 12, 2003, 17:32:19  
Job time : 448.085 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 17:17:59 ; Search time 3513.16 Seconds  
(without alignments)  
9028.139 Million cell updates/sec

Title: US-09-815-242-4228

Perfect score: 1305

Sequence: 1 agactaacctatagtagc.....atattatagctcgaagaga 1305

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estlin:\*
- 4: em\_estmu:\*
- 5: em\_estvrt:\*
- 6: em\_estvrt:\*
- 7: em\_estvrt:\*
- 8: em\_estvrt:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_est3:\*
- 12: gb\_est4:\*
- 13: gb\_est5:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_nam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rtd:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	486.6	37.3	2662	28	BH770770 LLMtag51
2	143.6	11.0	552	9	AI390536 mu83b02.y
3	128.4	9.8	721	12	BJ573090
4	124	9.5	1233	11	AY110834 2ea may5

5	118.6	9.1	583	14	CA782602
6	108.4	8.3	431	9	AA215138
7	99.8	7.6	579	13	BQ514303
8	99.2	7.6	532	13	BQ281519
9	89.4	6.9	395	9	AI465410
10	88.4	6.8	745	14	CD377507
11	84.6	6.5	171	28	BH234869
12	83.2	6.4	1101	29	CNS0039G
13	81.4	6.2	167	28	BH234795
14	79.6	6.1	653	13	BQ855888
15	78.8	6.0	501	14	N37957
16	78.8	6.0	613	10	BG299568
17	77.4	5.9	534	14	CA938405
18	76.8	5.9	631	28	BH861059
19	74.2	5.7	908	28	AZ548467
20	73.2	5.6	1359	28	BH770563
21	73.2	5.6	769	13	BQ802107
22	73.2	5.6	1010	14	CD102897
23	72.6	5.5	410	14	CD474342
24	71.6	5.5	841	14	CB996761
25	71.6	5.5	905	28	AZ550256
26	71.4	5.5	487	12	B1272296
27	71.4	5.4	538	13	BQ490035
28	70.6	5.4	552	14	CA742475
29	70.6	5.4	623	12	BJ463635
30	69.8	5.3	843	28	AZ551618
31	69.8	5.3	931	28	BH160372
32	69.2	5.3	412	12	BM402985
33	69.2	5.3	772	12	BI920601
34	69.2	5.3	849	28	AZ546009
35	68.8	5.2	912	28	AZ551092
36	68.4	5.1	745	28	AY079759
37	67.2	5.1	911	28	BH164417
38	66.6	5.0	512	10	AW979775
39	65.8	5.0	541	13	BQ627620
40	65.6	5.0	541	13	BQ627620
41	65.6	5.0	877	28	AZ531291
42	65.2	5.0	571	13	BU578174
43	64.2	4.9	721	10	BG646649
44	64.2	4.9	350	14	CA735279
45	63.6	4.9	356	14	H37009

#### ALIGNMENTS

RESULT 1  
BH770770  
LOCUS BH770770 2662 bp DNA linear GSS 01-MAY-2002  
DEFINITION LLMtag515 MG1363 Random Sequence Tag Library Lactococcus lactis  
subsp. cremoris genomic, genomic survey sequence.  
ACCESSION BH770770  
VERSION BH770770.1 GI:20373727  
KEYWORDS GSS.  
SOURCE Lactococcus lactis subsp. cremoris  
ORGANISM Lactococcus lactis subsp. cremoris  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Lactococcus.

REFERENCE 1 (bases 1 to 2662)  
AUTHORS Bolotin, A., Ehrlich, S. D. and Sorokin, A.  
TITLE Studies of genomes of dairy bacteria Lactococcus lactis  
JOURNAL Sci. Alimentis, (2002) In press  
COMMENT Contact: Sorokin A  
Genetique Microbienne

INRA  
CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France  
Tel: 33 1 34 65 25 16  
Fax: 33 1 34 65 25 21  
Email: sorokine@jouy.inra.fr  
best homologue in strain IL1403 is yphL (98%)  
Class: shotgun  
High quality sequence start: 30  
High quality sequence stop: 2634.



Uni-ZAP XR Vector: -5' adaptor sequence: 5' CAATCGGCACGAG  
3' -3' adaptor sequence: 5' CTCAGTTTCTTTTCTTTTCTTTT 3'  
BASE COUNT 220 a 63 c 94 g 174 t 1 others

Query Match 11.0% Score 143.6; DB 9; Length 552;  
Best Local Similarity 56.1%; Pred. No. 4.6e-19;  
Matches 316; Conservative 0; Mismatches 235; Indels 12; Gaps 2;

QY 136 GGTGAGTGGTTAAACACATGATTTCATATTATTATGATACAGGTGGTATTGAATTCGTGAT 195  
DB 1 GTTGATGAGCAGGAAGGAATTTATCTCTATAGACAGCGGGAATTTGAATAGAAAC 60  
QY 196 GCACATTCACAAACACAAATAGAGCGCAGCGCAAGAAATGCCATAGATGAAGCGGATTT 255  
DB 61 AAATCATTTCAAGAACAAATAGAAATTTCAAACTCAATAGCAATAGCAAGCAAAATTTA 120  
QY 256 ATTATTTTATGTTTAACTGCGTGAAGGATTGACACAAAGCGATGAAATGGTCTCAA 315  
DB 121 ATTATTTTATGTTTAACTGCGTGAAGGATTGACACAAAGCGATGAAATGGTCTCAA 180  
QY 316 ATTTTATACAAATCTAAACACCGTGTATTAGCGGTTAACAAAGTAGATTAATAGGAA 375  
DB 181 TTGCTTCGAAATCAAGCAAAAGTTTAAATCGCAGCTAATAAATAGAGGAAACAA 240  
QY 376 ATGCGTACAGACGTGATGATTCTATTATTAGGATTGGTGAACCGGTATCCAAATCA 435  
DB 241 TTTTCTT-----GATACTTCAATTTTACTAGTTTGAACATATTTTCCAAATTTCA 294  
QY 436 GGGTCACATGTTTAGTCTTGGTGGTCTTATAGTGCAGTTGTTCTCATTTTGGTGA 495  
DB 295 GCTATTCAGAGAGAGTGGTGGTATTATAGTGAACAAATAGAAATTTAGATTTT 354  
QY 496 GAGGAAGAAGATCTTATGTAAGATACAAATTCGACTATCCATTTATGGACGACCAAC 555  
DB 355 ACAAGA-----GAAGGAAATGAAGCATTTTAGTTGGCAATTTATAGGAAACCAAC 408  
QY 556 GTAGTAATCAAGTTTAGTAAATGCTATTTTAGTGAAGATCGCGTTATCTCTTAAT 615  
DB 409 GCTGGAATCTACTTTTAAATTAATAGCAATGAATCGCTCCATTTCTCTCCA 468  
QY 616 GTTCAGGCAACAGCAGACGCTATTGATACAGATATGTTATGATGACCAATAT 675  
DB 469 ATAGCAAGAACACAGCTGATTCAGTTCTTAGTTTATTAAATTTGCAAAATAGATNT 528  
QY 676 GTTTTAAATCGATCTGCTGGTAT 698  
DB 529 GAAATTATAGATCTGCTGGAAT 551

## RESULT 3

BJ573090/c 721 bp mRNA linear EST 18-DEC-2002  
LOCUS BJ573090 Ipomoea nil mixture of flower and flower bud Ipomoea nil  
DEFINITION cDNA clone jn22ml3 3', mRNA sequence.

ACCESSION BJ573090

VERSION BJ573090.1 GI:27254918

KEYWORDS EST.

SOURCE Ipomoea nil (Japanese morning glory)

## ORGANISM

Ipomoea nil  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;  
asterids; lamids; Solanales; Convolvulaceae; Ipomoea.

## REFERENCE

1 (bases 1 to 721)  
Hoshino,A., Seki,M., Shin-I,T., Carninci,P., Kamiya,A., Shiraki,T.,  
Nitasaka,E., Shinozaki,K., Hayashizaki,Y., Kohara,Y. and Iida,S.  
ESTs of Japanese morning glory

## TITLE

Unpublished

## COMMENT

Contact: Tadasu Shin-I  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856

Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.  
Location/Qualifiers  
1..721  
/organism="Ipomoea nil"  
/mol\_type="mRNA"  
/cultivar="Tokyo-kokei standard"  
/db\_xref="taxon:35883"  
/clone="j22ml3"  
/tissue\_type="mixture of flower and flower bud"  
/clone\_lib="Ipomoea nil mixture of flower and flower bud"

## FEATURES

source

BASE COUNT 186 a 184 c 136 g 215 t  
ORIGIN

Query Match 9.8% Score 128.4; DB 12; Length 721;  
Best Local Similarity 52.8%; Pred. No. 6.2e-16;  
Matches 378; Conservative 0; Mismatches 321; Indels 17; Gaps 4;  
QY 587 TAGGTGAGATCGCTTATCGTTTCTAATGTTGCGAGGCAACGAGAGCGCTATTGATA 646  
DB 719 TTGGGAGAGACAGACAATTTGTTAGTCCGTCAGTGGAACTACTCTGATGCAATTTGATA 660  
QY 647 CAGAGTATA---GTTATGATGCAAGATTTATGTTTAAATCGATCTCTGTTGATCGCTA 703  
DB 659 CAGAGTTTACTGGATCAGATGACAGAAATATCGTCTCATTTGATCTGCGAGGAATAGGA 600  
QY 704 AAAAGGAAAGATATATCAATC-----AAGTGAAGAAATTTCAAGTATTAAAGAGCTT 754  
DB 599 AAAGGCAACTCTGGCTTCATCTGTAGCATAACAGAGCTTTATCTGTAATCAAGCAT 540  
QY 755 TAAAGCGATTCAGCTTCAATGTTGTTTAGTGGTTATTGATGCAAGAACAGGCAATTA 814  
DB 539 TCGTGCAATAGCTGCTCAGATGTTGGCTCTTGTAAATTTAGGCTATGCTCTTGCATCA 480  
QY 815 TTGAACAGATAAAGCTGTTCAGGATATGCACATGAACAGGTAAGAGAGTCTGATTTG 874  
DB 479 CTGAGCAGGATTCGAAGATTCGAGAAAGATAGAGAGGAAAGGTAAGGTTGCTCTGTTG 420  
QY 875 TCGTAATAAATGGGATCTGTGGAAAGATAGTAA--AAGATGAAGAAATTTGAAGA 932  
DB 419 TTGTAATAAGTGGGACACAAATTCACAAACCAATCACTGCAACTTACTATAGCA 360  
QY 933 TGAAGTACGTAAGAATTTCCAAATTTTAGATATTGCAAAATTTGCTTTGCTGCTGCTAA 992  
DB 359 AGATGTTAGGGGGAAGCTCCGTTGTTGGTTGGCCACCTATTGTTGATTCAACTGCAAT 300  
QY 993 AGAACGCACAGATTACGTACATTTATCCCTTACATTAATTAAGCAAGTGAAACCAATAA 1052  
DB 299 AGCAGGCATACGCTCGACAAAATTTATTTGTTACTGCTGCTGTTGAAAAAGAGATC 240  
QY 1053 AAAAGCTGTTCAAAGTTCAACTTTAAATGAAGTCTTACTGATGCAATTTCCATGAACCC 1112  
DB 239 AGAAGATTAAACACGCGCAATTCGAATCAAGTGGTAGGGAGGAGCTGCGGTTCAATC 180  
QY 1113 TACACCAACAGACAAAGGT---AGACGTTTGAATGCTCTTTTATGCAACACAAAGTTCTAT 1169  
DB 179 ACCTCCAGGACTAGAGGTGGGAGAGGGGCTGCTATATTATTGCACTCAGGCTGCTAT 120  
QY 1170 AGAACCCAGCAGATTGTTGTTATTTGTTAATGATGATAGAATTAATGCAATTTTCTTATAA 1229  
DB 119 TAGACCCACTACTTTTGTCTTTGTCATGATGCAAAACTTTTCTCCGAGACATACCG 60  
QY 1230 ACCTATTTTAGAATCAAAATCCGTCGCGCTTTTGGCTTTTGAAGGTACACCAATTC 1285  
DB 59 ACGGTACATGGAAAAAGCAACTGAGATCAAGTGCAGGGGTTCTCGGCGCACACCTATTC 4

## RESULT 4

AV110834

LOCUS AV110834

DEFINITION Zea mays CL915\_1 mRNA sequence.

ACCESSION AV110834

VERSION AV110834.1 GI:21215424

AV110834 1233 bp mRNA linear

HTC 17-OCT-2002









```

Best Local Similarity 53.3%: Pred. NO. 5.5e-10; Mismatches 218; Indels 6; Gaps 2;
Matches 256; Conservative 0;

QY 753 TTTAAAGCGATTGAACGTTCAAAATGTTTGTAGTGGTATTATGTCAGAACAGGCAT 812
Db 48 TTTTCTAGCAATTCGCGCTCTGATGTTGCTCCCTTGTATCTAGGCAATGGCGTGT 107
QY 813 TATTGAACAAGATAAAGCTGTTGTCAGGATATGACATGAACAAGTAAGCAAGTCGTGAT 872
Db 108 TACCGACCAAGATTATATAAATTCACAGAGGATTGCAAGAAGAAAGCTTGTGTCAT 167
QY 873 TGTCTTAATTAATGGATCTGTGGAAGAAGATA---GTAAACGATGACAAATTTGA 929
Db 168 TGTGTGAACAATGGGATGATCTATCCCTAACACAGACCATGAGATACACACATATGA 227
QY 930 AGATGAATACGTACGTAACGAATTTTATGATATGACAAATTTGTTGTGTCGC 989
Db 228 ACAAGATGTAATAGAAAGCTTCCCATCTGTTGGCACCCTATTGTCTACTGTTCTGC 287
QY 990 TAAAGAAGCGACGAAGATTAGCTATTTTCCCTTACATTAATGAAGCAAGTGAACCA 1049
Db 288 GACTAATGCGACCAAGTGTCAAAAGATATTTCTGCTGCTGCTTGTGATGAATGAAG 347
QY 1050 TAAAAACGTTTCAAGTTCAACTTTAATGAAGTGTGTTACTGATGCAATTTCCATGAA 1109
Db 348 GTCTATAGACTCGGCACCTCCATTTCTTAATCAAGTTATAGAGAGCTATAGCATTCAA 407
QY 1110 CCCTACACCAACAGACAAAGT---AGACGTTTGAATGCTTTTATGCAACCAAGTTGC 1166
Db 408 ACCACCAACGAAGAGGTTGCAAAAGAGGCTGCTGCTCTATATACACACAGGCTGC 467
QY 1167 TATGACACCAACGATTTGTTGTTTATGTTTATGATGATGATGATGATGATGATGAT 1226
Db 468 CATTGGTCCCAACCAATTCGTTTATTTTGTAAATGATGGAAGCTCTTCCCTGATACATA 527

RESULT 9
AI465410/c
LOCUS
DEFINITION
mu83b02.x1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
clone IMAGE:652107 3', similar to SW:PHC-BACSU P50743 HYPOHETICAL
48.8 KD GTP-BINDING PROTEIN IN CMK-GPSA INTERGENIC REGION. ; mRNA
sequence.
ACCESSION
AI465410
VERSION
AI465410.1 GI:4319440
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 395)
AUTHORS
Marras M., Hillier L., Kucaba T., Martin J., Beck C., Wylie T.,
Underwood K., Stepien M., Theising B., Allen M., Bowers Y., Person
B., Swaller T., Gibbons M., Pape D., Harvey N., Schurk R., Ritter
E., Kohn S., Shin T., Jackson Y., Cardenas M., McCann R.,
Waterson R. and Wilson R.
The WashU-NCI Mouse EST Project 1999
Unpublished
TITLE
Marras M/NCI Mouse EST project 1999
JOURNAL
Washington University School of Medicine
COMMENT
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 276.
Location/Qualifiers
1..395
/organism="Mus musculus"

/mot_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:652107"
/tissue_type="melanoma"
/dev_stage="M2 cells"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse melanoma (#937312)"
/Note="Organ: skin; Vector: pBluescript SK-; Site: 1: EcoRI
Site: 2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. From M2 cells, a highly metastatic derivative of the
K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
Uni-ZAP XR vector; -5' adaptor sequence: 5' GAATTCGCCAGG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."

BASE COUNT 140 a 41 c 47 g 167 t
ORIGIN
Query Match 6.9% Score 89.4; DB 9; Length 395;
Best Local Similarity 55.3%; Pred. No. 5.3e-08;
Matches 194; Conservative 0; Mismatches 156; Indels 1; Gaps 1;

QY 955 TTTTATGATTTATGACAAATGCTTTTGTCTGCTTAAGACGACAGATGATGATGATGATGAT 1014
Db 364 TTTTATGCTGAGCCCAATTTGTTTATTTTACG-ACACAGGAGATGATGATTTCTTAAA 306
QY 1015 TTTATCCCTTACATTAATGAAGCAAGTGAACCAATTAAGCAAGTGAACCAAGTGAAC 1074
Db 305 TTAAGAGAACCAATTTTGAAGATCAAAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAA 246
QY 1075 TTAATGAAGTCTGTTACTGATGCAATTTCCATGACCTACACACACACACACACACAC 1134
Db 245 CTTAATCAATTCATTTAGATATTCATGTTTACACCTCCCAAGTGTCTGCTGGAAGA 186
QY 1135 CTTTGAATGCTTTTATGCAACACAAAGTGTGCTATGACACCAAGTGTGCTGATGAT 1194
Db 185 AAATTAATTAATTAATTTGCTCAACAAATTAATTAATTAATTAATTAATTAATTAATTA 126
QY 1195 GTTAATGATGATGATTAATGCTATTTTCTTATTAACGCTATTTATGAGATCAATTCGT 1254
Db 125 GTTAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 66
QY 1255 GCGCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1305
Db 65 AACTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 15

RESULT 10
CD377507
LOCUS
DEFINITION
PTM02676 Phaeodactylum tricornutum Uni-zap XR Phaeodactylum
tricornutum cDNA 5', mRNA sequence.
ACCESSION
CD377507.1 GI:31253121
VERSION
EST.
KEYWORDS
Phaeodactylum tricornutum
SOURCE
Phaeodactylum tricornutum
ORGANISM
Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae;
Bacillariophycidae; Naviculales; Phaeodactylaceae; Phaeodactylum.
1 (bases 1 to 745)
REFERENCE
Scala S., Carls N., Falciatore A., Chiusano M.L. and Bowler C.
Genome properties of the diatom Phaeodactylum tricornutum
Plant Physiol. 129 (3), 993-1002 (2002)
2211123
12114555
Contact: Bowler C
Laboratory of Molecular Plant Biology
Stazione Zoologica 'Anton Dohrn'
Villa Comunale, I-80121, Napoli, Italy
Tel: 39 081 583 3268/3211
Fax: 39 081 764 1355
Email: chris@lephs.zsn.it
Diatom EST Database(http://avesthagen.szbowler.com)
Seq primer: 13 backward
POLYA=No.

```

FEATURES  
source Location/Qualifiers  
1. 745  
/organism="Phaeodactylum tricornutum"  
/mol\_type="mRNA"  
/db\_xref="taxon:2850"  
/cell\_line="CCMP632"  
/clone\_lib="Phaeodactylum tricornutum Uni-Zap XR"  
/note="vector: Uni-Zap XR vector; Site\_1: Eco RI; Site\_2: Xho I"  
Xho I 179 a 182 g 182 g 198 t 4 others

BASE COUNT  
ORIGIN 179 a 182 g 182 g 198 t 4 others

Query Match 6.8%; Score 88.4; DB 14; Length 745;  
Best Local Similarity 51.5%; Pred. No. 9.1e-08;  
Matches 281; Conservative 0; Mismatches 256; Indels 9; Gaps 3;

QY 749 GAGCTTTAAAGCGATTCAAGTTCAATGTTGTTTGTAGTGTATTGATGCGAGACAAG 808  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 40 GAGCACTCAGCGGATCGCGTCCGATGCGTCTTCTTATCTGGATGCCACTTCCG 99  
QY 809 GCATTATTGAACAAGATAAAGTGTGTCAGGATATGCACATGACAAAGTAAGCAGTCG 868  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 100 GTGTAGCGCAACAAGATCGTCTCTTGGCGCAGAAATTTGCCGATGATGACGCCATCG 159  
QY 869 TCATTGTCTGAATAAATGGGATCTGTG--GAAAAGATAGTAAACGATGAAGAAAT 925  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 160 TGATCGTTGCACAAATGGGATCTGTCTGTTGATAGGATTCACACAGTACGACAAGT 219  
QY 926 TTGAAGTGAAGTACGTAAGAAATTCCAATTTTGTAGATTATGACAAATTTGCTTTGTGT 985  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 220 CGGTCCCAATATCTTGGAGAAGATTTGCCGATGATCTGTTGGCCCTATCTTATTATATCT 279  
QY 986 CTGCTAAAGACGCAAGATTACGTACATATTTCCTTACATTAATGAAGCAAGTGAAA 1045  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 280 CGGCTGCCACTGGGCAACGCTGTGGCAAGATATACAGCGCATTTGCGGTGCCATCGAAG 339  
QY 1046 ACCATAAAGCGTTTCAAGTTCACTTTAATGATCGTTACTGATGAATTTTCCA 1105  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 340 CTCATCTGAACGATAGACGCGTATCTTCACGAAGTTTGGAGATGCTATTATTGT 399  
QY 1106 TGAACCCCTACCAACA--GACAAAGGTAGAGTTTGAATGTCTTTTATGCAACACAAG 1162  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 400 GCGACCCACCTCCGACACGCGGTCACAGCGGTCACAGCGGAAGATATACTACTGCAACCAAG 459  
QY 1163 TTGCTATAGAACCCGACATTTGTTGTTATTTGTTATGATGTAGATTAATGCAATTTT 1222  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 460 TGAGTACCGGACCACTACCGTCTGTTTCTGCAATGATCCCAACTGGTCAAGACA 519  
QY 1223 CTTATAAGCGTATTAGAGNATCAATCCGTCGCCCTTT--TGTGTTTGAAGTAGAC 1279  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 520 ATTACGCGGTACTTGGATCGTAAGTTCCGTAATCACTGGATGGATTTGAAGCGACTC 579  
QY 1280 CAATTC 1285  
|||||  
Db 580 CCATTC 585

RESULT 11  
LOCUS BH234869  
DEFINITION PE.04.x E Spiroplasma kunkelii genomic clone PE.04.x, genomic survey sequence.  
ACCESSION BH234869  
VERSION BH234869.1 GI:18030337  
KEYWORDS GSS.  
SOURCE Spiroplasma kunkelii  
ORGANISM Spiroplasma kunkelii  
Bacteria; Firmicutes; Mollicutes; Entomoplasmatales; Spiroplasmataceae; Spiroplasma.  
REFERENCE Hogenhout.S.A.  
AUTHORS I (bases 1 to 171)  
TITLE Genomic sequences from Spiroplasma kunkelii strain M2  
JOURNAL Unpublished

COMMENT Contact: Hogenhout SA  
Department of Entomology  
The Ohio State University-OARDC  
120 Thorne Hall, 1680 Madison Avenue, Wooster, OH 44691, USA  
Tel: 330 263 3730  
Fax: 330 263 3686  
Email: hogenhout.1@osu.edu  
Class: EcoRI fragments.  
Location/Qualifiers  
1. 171  
/organism="Spiroplasma kunkelii"  
/mol\_type="genomic DNA"  
/strain="M2"  
/db\_xref="taxon:47834"  
/clone\_lib="PE.04.x"  
/clone\_lib="E"  
59 a 26 c 26 g 60 t

BASE COUNT  
ORIGIN 59 a 26 c 26 g 60 t

Query Match 6.5%; Score 84.6; DB 28; Length 171;  
Best Local Similarity 68.4%; Pred. No. 4.7e-07;  
Matches 117; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 525 AATTCGACTATCCATTATTGGACGACCAACAGCTAGGTAAATCAAGTTTAAATGCTAT 584  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 171 AATTCATTTTCTTTAATCGGTAAACCAATGTTGGAAAATCTCTTTACAAATGCTAT 112  
QY 585 TTTAGTGGAAGATCCGTTTATGCTTTCTAATGTTGAGGACCAACGAGAGCGCTATTGA 644  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 111 TTTAGTGGAAGAACAGTATTGTTGATCACCATTGTCAGGAACACTACAGATTCAATTGA 52  
QY 645 TACAGAGTATAGTATGATGACCAAGATTATGTTTAAATCATCTACTGCTGG 695  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 51 TACATCATTTAAACGGAATATCACTTTATCTGCTGATGCTACTGCTGG 1

RESULT 12  
LOCUS CNS0039G/c  
DEFINITION Drosophila melanogaster genome survey sequence JET3 end of BAC # BAC08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL063921  
VERSION AL063921.1 GI:4941778  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
Location/Qualifiers  
1. 1101

FEATURES  
source 1. 1101

```

/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
BASE COUNT      201 a      64 c      131 g      202 t      503 others
ORIGIN

Query Match      6.4%   Score 83.2;   DB 29;   Length 1101;
Best Local Similarity 19.6%   Pred. No. 1.1e-06;
Matches 130;   Conservative 271;   Mismatches 258;   Indels 3;   Gaps 1;

QY 441 ACATGGTTAGCTTGGTGGACGTTGTAGATGCGATGTTCTTCATTTTGTGGAAGAGGA 500
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1099 ARRGGDTWDTRDKDDWMTKWTWKDKDRDRRWDGADRWAMDDGAGTWTATWMMW 1040
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 501 AGAAGATCCTTATGATGAAGATCAACAATTCGACTATCCATTTATTCGACCAACCAAGTAGG 560
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1039 WMMWATWDTWKKMMWATAAKTDTATWMTAWRADWAGRDGAGKRDRAATDADGAG 980
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 561 TAATCAAGTTTAAATGCTATTTTAGTGAAGATCGCTTATCTTCTTCTATGTTGC 620
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 979 RRGGRKKDKKDDDDKGGKKKAAKAAKAWATKMWDDMDKMDKWDGAKDRAD 920
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 621 AGGCACACGAGACGCTATTGATACAGATATAGTTATGATGACACAGATATGTTTT 680
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 919 DDDGAGDKDDGKDDDDTGGTDDDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDD 860
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 681 AATCGATCTGCTGATGGTAAAGAAAGAAAGATATGATGAATCAACTGAGAATATTC 740
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 859 ADADWTTWDAADWDWADWDWADWADWADWADWADWADWADWADWADWADWADWADWAD 800
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 741 AGTATTAAAGCTTTAAAGCGATTGAACGTTCAAAATGTTGTTTATGTTGTTATGATGC 800
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 799 KRADDKDAADDDDAATWTTTWTTR---DTDDKWKTKTDTWTRMAADRTWDRDDDD 743
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 801 AGAACAGGCTATTATGAACAGATAAAGCTGTCGACGATATGACGATGACGATGACGATGAA 860
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 742 DRAGTAGKRRRTWKRRKRRDRDADAADDTARDRRRRRGGDADAGKGGKTKGRKRR 683
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 861 AGCAGTCTGATCTGCTAATAAATGGCTACTGCGAANAAGATAGTAAACGATGAA 920
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 682 RDRATWDTDAWADAANWTTTDTDDMDKDRRRKGARRRRRTTARAADWMTWAKWD 623
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 921 GAAATTTGAAGATGAAGTACGTAAGAAATTCCAATTTTATGATGACCAATGCTTT 980
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 622 WAKDWKTRADRWDAADTWTDAKADRWAKARARRRDRARARARADRRWTTKGKTT 563
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 981 TGTCTGCTGCTAAAGACGACAGATTACGATATTTCCCTTACATTATGAGGCAAG 1040
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 562 TATWTTWAARAANWAWANATTTATWTTTWTWTTTWTWTTTWTWTTTAAWNAWTA 503
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 1041 TGAACACATAAAGAGCTGTTCAAGTTCACATTTAAATGAAGTCGTTACTGATGCAAT 1100
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 502 TAAATTAATAAANAANAANAATTTTTTTTTTTTAAWNAWNAWNAWNAWNAWNAWNAWNA 443
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 1101 TT 1102
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 TT 441

RESULT 13
BH234795
LOCUS
DEFINITION
BH234795
ACCESSION
BH234795
VERSION
BH234795.1
KEYWORDS
GSS.
SOURCE
Spiroplasma kunkelii
ORGANISM
Bacteria; Firmicutes; Mollicutes; Entomoplasmatales;

/organism="Spiroplasma kunkelii"
/mol_type="genomic DNA"
/strain="M2"
/db_xref="taxon:47834"
/clone="MEAA_B03.y"
/clone_lib="E"
BASE COUNT      58 a      26 c      26 g      57 t
ORIGIN

Query Match      6.2%   Score 81.4;   DB 28;   Length 167;
Best Local Similarity 68.7%   Pred. No. 2.1e-06;
Matches 112;   Conservative 0;   Mismatches 51;   Indels 0;   Gaps 0;

QY 533 TATCATTTATGGAGCAGCAACCTAGTAATCAAGTTTAGTAATGCTATTTAGGTG 592
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 TTTCTTTAATCGTTAAACCAATGTTGCAAAATCTCTTTTACAAATGCTATTTAGGTG 64
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 593 AGATCGCGTTTCTGTTTCTTANTGTTCCAGGACAGAGACGCTATTTGATACAGAGT 652
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 AAGAAGCAGTATTTGTTATCACCAGTTCAGGACCACTACAGATTCAATTGATACATCAT 124
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 653 ATAGTTATGATGACGACAGATTTATTTTAAATCGATCTGCTGG 695
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 TTAAACGAATAAATCACTTTTATATCTGTGATTGATCTGCTGG 167
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
BQ55888
LOCUS
DEFINITION
BQ55888
ACCESSION
BQ55888
VERSION
BQ55888.1
KEYWORDS
EST.
SOURCE
Lactuca sativa
ORGANISM
Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichoraceae; Lactuca.
REFERENCE
1 (bases 1 to 653)
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevallier, P., Ziegler, J., Ellison,
P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Composite Genome Project
http://compgenomics.ucdavis.edu/
Unpublished
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmoreveg@mail.ucdavis.edu]
singleton; see http://cgdb.ucdavis.edu/ for details.
Plate: QGB28 row: A column: 05.

BQ55888      653 bp      mRNA      linear      EST 14-AUG-2002
QGB28A05.yg.ab1 OG-ABCDI lettuce salinas Lactuca sativa cDNA clone
QGB28A05, mRNA sequence.
BQ55888
BQ55888.1      GI:22241353
EST.
Lactuca sativa
Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichoraceae; Lactuca.
REFERENCE
1 (bases 1 to 653)
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevallier, P., Ziegler, J., Ellison,
P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Composite Genome Project
http://compgenomics.ucdavis.edu/
Unpublished
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmoreveg@mail.ucdavis.edu]
singleton; see http://cgdb.ucdavis.edu/ for details.
Plate: QGB28 row: A column: 05.
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2003, 17:17:09 ; Search time 5327.98 Seconds  
(without alignments)  
10020.134 Million cell updates/sec

Title: us-09-815-242-4228

Perfect score: 1305

Sequence: 1 agcactaacctatagtagc.....atattatagctgaaagaga 1305

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0  
Maximum Match 1000  
Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sv.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rtd.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgtgo\_hum.\*  
40: em\_hgtgo\_mus.\*  
41: em\_hgtgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	1292.2	99.0	301550	1	AP003134	AP003134 Staphyloc
C 2	1292.2	99.0	346900	1	AP003362	AP003362 Staphyloc
C 3	1290.6	98.9	1308	6	AX622668	AX622668 Sequence
C 4	1290.6	98.9	333750	1	AP004827	AP004827 Staphyloc
C 5	919.4	70.5	3259	1	AF270032	AF270032 Staphyloc
C 6	919.4	70.5	3259	6	AX145350	AX145350 Sequence
C 7	919.4	70.5	300892	1	AE016747	AE016747 Staphyloc
C 8	733.8	56.2	304680	1	AE017002	AE017002 Bacillus
C 9	727.4	55.7	290117	1	AE017028	AE017028 Bacillus
C 10	683.2	52.4	347050	1	AL591981	AL591981 Listeria
C 11	683.2	52.4	349980	6	AX641670	AX641670 Sequence
C 12	676.8	51.9	313450	1	AL596170	AL596170 Listeria
C 13	676.8	51.9	319630	6	AX413016	AX413016 Sequence
C 14	676.8	51.9	349980	6	AX417046	AX417046 Sequence
C 15	667.6	51.2	960	6	AX144037	AX144037 Sequence
C 16	656.8	50.3	300050	1	AP004599	AP004599 Oceanobac
C 17	647.4	49.6	24887	1	BACSERA	L47648 Bacillus su
C 18	647.4	49.6	213680	1	BSUB0012	Z99115 Bacillus su
C 19	634.8	48.6	302173	1	AE016951	AE016951 Enterococ
C 20	633	48.5	300550	1	AP001512	AP001512 Bacillus
C 21	600	46.0	1311	6	AX607165	AX607165 Sequence
C 22	600	46.0	20501	1	AE014265	AE014265 Streptoco
C 23	600	46.0	44145	6	AX602195	AX602195 Sequence
C 24	600	46.0	174050	1	SAG766852	AL766852 Streptoco
C 25	599.4	45.9	1311	6	AX433884	AX433884 Sequence
C 26	582.4	44.6	11071	1	AE006309	AE006309 Lactococc
C 27	552	42.3	52276	1	AE014141	AE014141 Streptoco
C 28	552	42.3	151947	2	SPNEU1902	AL449524 Streptoco
C 29	552	42.3	323825	1	AP005146	AP005146 Streptoco
C 30	550.4	42.2	1308	6	AX570314	AX570314 Sequence
C 31	550.4	42.2	5066	6	AR218847	AR218847 Sequence
C 32	550.4	42.2	5066	6	BD003759	BD003759 Polynucle
C 33	550.4	42.2	10310	1	AE007464	AE007464 Streptoco
C 34	550.4	42.2	349980	6	AX571764	AX571764 Sequence
C 35	550.4	42.2	349980	6	AX571765	AX571765 Sequence
C 36	549.6	42.1	12434	1	AE006498	AE006498 Streptoco
C 37	548.8	42.1	12370	1	AE009978	AE009978 Streptoco
C 38	548.8	42.1	12540	1	AE008523	AE008523 Streptoco
C 39	544.2	41.7	302050	1	AL935257	AL935257 Lactobaci
C 40	531.2	40.7	3737	1	AB016077	AB016077 Streptoco
C 41	531.2	40.7	13860	1	AE015016	AE015016 Streptoco
C 42	486.2	37.3	301278	1	AE015939	AE015939 Clostridi
C 43	476	36.5	296750	1	AP003191	AP003191 Clostridi
C 44	438.4	33.6	10861	1	AE007680	AE007680 Clostridi
C 45	423.2	32.4	10691	1	AE010530	AE010530 Fusobacte

# ALIGNMENTS

RESULT 1  
AP003134/c  
LOCUS AP003134 301550 bp DNA linear BCT 24-APR-2003  
DEFINITION Staphylococcus aureus subsp. aureus N315 genomic DNA, complete  
genome, section 6/10.  
ACCESSION AP003134 BA000018  
VERSION AP003134.2 GI:14349226  
KEYWORDS  
SOURCE Staphylococcus aureus subsp. aureus N315  
ORGANISM Staphylococcus aureus subsp. aureus N315  
Bacteria; Firmicutes; Bacilliales; Staphylococcus.  
REFERENCE  
AUTHORS Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,  
Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Iino, J., Ito, T., Kanamori, M.,  
Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A.,



Mizutani-Ui, Y., Takahashi, N.K., Sawano, T., Inoue, R., Kaito, C.,  
 Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J.,  
 Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C.,  
 Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., and Hiramatsu, K.  
 Whole genome sequencing of methicillin-resistant *Staphylococcus*  
*aureus*  
 Lancet 357 (9264), 1225-1240 (2001)  
 21311952  
 11418146  
 2 (bases 1 to 301550)  
 Director-General, Biotechnology Center, Aoki, K., Oguchi, A.,  
 Hosoyama, A., Nagai, Y., Kuroda, M., Hiramatsu, K. and Kikuchi, H.  
 Direct Submission  
 Submitted (30-JAN-2001) Director-General, Biotechnology Center,  
 National Institute of Technology and Evaluation, Biotechnology  
 Center, 2-chome 49-10 Nishihara, Shibuya-Ku, Tokyo 151-0066, Japan  
 (E-mail: biocente.go.jp, URL: http://www.bio.nite.go.jp/  
 Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)  
 On Jun 12, 2001 this sequence version replaced gi:13701258.  
 Location/Qualifiers  
 1. 301550  
 /organism="Staphylococcus aureus subsp. aureus N315"  
 /mol\_type="genomic DNA"  
 /strain="N315"  
 /sub\_species="aureus N315"  
 /db\_xref="taxon:158879"  
 /complement(249. .947)  
 /gene="SA1293"  
 /complement(249. .947)  
 /gene="SA1293"  
 /note="ORFID:SA1293"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAB42553.1"  
 /db\_xref="GI:13701259"  
 /translation="MSFTNIIYFVILVPLHQAQHVKNSEKYSQVRSCTKGR  
 EVALLHANGITDVEVGEGLFDHDPKRVVLSLSPANDRPSVAGTAIAAEVHG  
 AADQHGQVFLFRALVFPANLGSLSYKIIHLGIIITLTAIGSAFGSTALWAGLH  
 SLAVFSITLPEVFPDASSRANKQITALNIVNEKEYHARKVLSAAAMTYVNAVAV  
 AEIVRIILIARSSD"  
 /complement(1001. .1588)  
 /gene="SA1294"  
 /complement(1001. .1588)  
 /gene="SA1294"  
 /note="ORFID:SA1294"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAB42554.1"  
 /db\_xref="GI:13701260"  
 /translation="MTINTFWQYTLQYRSWLMMLILCNILGNIYGIWYGRQSLTPW  
 HFRIYVPSPTAILYVISLILQKQNSIIEALAFVTLFKYGIWAVIMNLIIFIQ  
 GDITVGLVMSHISIMAVQAIYVFPKRSNIGISVANWVFLNDVIDYFHLQFPY  
 DFITTHWQIGVLSCLSVFGLLILNKLKCK"  
 /complement(1578. .2153)  
 /gene="SA1295"  
 /complement(1578. .2153)  
 /gene="SA1295"  
 /note="ORFID:SA1295"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAB42555.1"  
 /db\_xref="GI:13701261"  
 /translation="MSETLNQIKESFIEVLLFQYRFRKRIANWNLNIVNEAKLANI  
 HFVTRLNHETAEVGSASHASAOFTARNIKLMTNEIFDIYANHCAPDIQIHFA  
 NVSKREORDLIVAQUTESPQYTLHDLNSMADRKHALLIDYLLHNIDLSLQNM  
 EKQRFQTLQILNKLKLNKHNQFEDLADD"  
 /complement(2167. .3411)  
 /gene="SA1296"  
 /complement(2167. .3411)  
 /gene="SA1296"

/note="ORFID:SA1296"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="BAB42556.1"  
 /db\_xref="GI:13701262"  
 /translation="MEDIYKLIDIDINLQKLENLDSRVNEAITTNDNDALFILGETLYN  
 FGLPQGLEVFRVLYHKIPDESELLYFIEGLMSENQDEALEYLSYVPSPEKMLLE  
 ADLYQOINMVEAIDKQEALELEPNDFIHFALAEMLYDQYLRYATSEYETVLETG  
 EYQVGNVLFNRADCKQASGNTSDAIRLYDEINEMEDSEYLLAKAISYDNQITOC  
 AIKMTTLLSKDPDYIQGYLYLQSLYENENKYPDAIETGKEGLRSQFYKELMTTTC  
 LEIGHGANEGVLLLKQALEVNAYQEPILLISDLYRNEEDYEAIIELIYVDEEDLO  
 PTFMHLAFAYGOBERKDEAKOHEFFELAYPTMTNTIDFNSDYFYFYLEIGONEQAITLL  
 NOLLEPSNENHDELSRLQS"  
 /complement(3418. .4716)  
 /gene="arob"  
 /complement(3418. .4716)  
 /gene="arob"  
 /note="ORFID:SA1297"  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-phosphoshikimate 1-carboxyvinyltransferase"  
 /protein\_id="BAB42557.1"  
 /db\_xref="GI:13701263"  
 /translation="MSEQIIDISGPKGIEVPGDKSMTHRAIMLASLAEGETSNLYK  
 PLLGECRRRTMDIFRLGVDIKEDOKLVNSPGYKAKTPOVLYTGNSTVTRLLA  
 GLSLGIGIESVLSIGARKPRDVRRLKLDANKEGIEDNTYPLIKRPSVINGIN  
 YONEVASAQVKSAILFASLFSNNTVIKELDSRNHTTFRHFNPIEAERLSITTT  
 PDAIQHIKPADHFVPGDISIAAFFVAAALITPESDVTIHWGINSRPSGIIDIVERKMG  
 GNIQLOFTNOTTGAETPISIRIYQTPMLOPITIEGELVPAKIDELPALLLCIOAVGTST  
 IKDAELKVKETNRIDTADMLNLGFELOPTNDGLIHPSEFKTNATVDSLTDRHIG  
 MMLAVALLSEPVKIQPDANVNSPFGFLPKLLENEG"  
 /complement(4726. .5790)  
 /gene="arob"  
 /complement(4726. .5790)  
 /gene="arob"  
 /note="ORFID:SA1298"  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-dehydroquinate synthase"  
 /protein\_id="BAB42558.1"  
 /db\_xref="GI:13701264"  
 /translation="MKLQTPSPNNTPIFEVHGAIDHISTYIDQFDSFILIDEHVNO  
 YFADKFDIILSYENHVKVIIPAGEKTKTEQYQETLEYILSHHVTNTAITAVGGGAT  
 GDFAGFAALTLCRGVHFIOVPTTILAHDSVGVKGVKINSKQKNLIGAFYRPTAVYD  
 LDFLTLPFOILSGYAEVYKHALNGESTTOEIOHFKDREILQSLGAMDKYIAGKI  
 ETKLDIVVADEKEQGVKFLNLGHTGHAVEYNHKLHAGHAVMIGIIVQFIVANLLEN  
 SNHDIQHYINKLKLGYPLETIDIDFETIYQYMLSDKKNDKQGVQWVLIHFGDIVV  
 QHIDQTLQRAACEQLATYFK"  
 /complement(5816. .6982)  
 /gene="arob"  
 /complement(5816. .6982)  
 /gene="arob"  
 /note="ORFID:SA1299"  
 /codon\_start=1  
 /transl\_table=11  
 /product="chorismate synthase"  
 /protein\_id="BAB42559.1"  
 /db\_xref="GI:13701265"  
 /translation="MRYLTSGESGHPQLTVIVGIPANLEIKVEDINKEMFKROGGYV  
 RGRMQIEKDTVEIVSGVRNGYTLGSPITWVVTNDDFTHWKIMGAAPISSEENMK  
 RTITKPRGHADLVGGHKYHNRDLRNVRSSARETAARVAYGALCKVYLQQLDIDY  
 SRVVEIGIKDKDFYDSETEFKANLDRNDVRVIDDSIAQMRDKIDEAKNEGDSIGGVY  
 QVYVENMPVGVSVHYDRKLDGKIAQGVYSINAFKGVSGFSGEKAEEKPGSEIQDEI  
 LYNSEICYRGSNHLLGEGGMSNGPIIVNGVMKPIPTLYKPLNSVDINTKEDFKAT  
 IERSDCAVPAASIVCEHVVAEIAKALLEEFSQNSHIEQLQQLQIADRLQVNEP"  
 /complement(7368. .7568)  
 /gene="SA1300"  
 /complement(7368. .7568)  
 /gene="SA1300"  
 /note="ORFID:SA1300"  
 /codon\_start=1  
 /transl\_table=11

		/product="hypothetical protein"	
		/protein_id="BAB42560.1"	
		/db_xref="GI:13701266"	
		/translation="MKNIISIIILGILMLPLKMLLMLLYGAIFLQPLRNPITKIIFILTLI YIFVLVELIIFLKSRYKNSA"	
gene		complement(7597..7722)	
CDS		/gene="SAS045"	
		complement(7597..7722)	
		/gene="SAS045"	
		/note="ORFID:SAS045"	
		/codon_start=1	
		/transl_table=11	
		/product="hypothetical protein"	
		/protein_id="BAB42561.1"	
		/db_xref="GI:13701267"	
		/translation="MWLVIMIHKKARFLCCSKLLEKFSYFLVTRIKRGRIFIK"	
gene		complement(7777..8226)	
CDS		/gene="ndk"	
		complement(7777..8226)	
		/gene="ndk"	
		/note="ORFID:SA1301"	
		/codon_start=1	
		/transl_table=11	
		/product="nucleoside diphosphate kinase"	
		/protein_id="BAB42562.1"	
		/db_xref="GI:13701268"	
		/translation="WERTFLMIKPDVORNIGEVISRIERKGLKLVGKGLMQVPMEL AETHYGEHQGPYNDLITSPITSAPVAMVVEGEDAVNYSRHIIIGSTNPSEASPGSIR GDLGLTVGRNIIHGSDLSLESAEREINLWENENETTSVSPDANLVE"	
gene		complement(8318..9277)	
CDS		/gene="gerCC"	
		complement(8318..9277)	
		/gene="gerCC"	
		/note="ORFID:SA1302"	
		/codon_start=1	
		/transl_table=11	
		/product="heptaprenyl diphosphate synthase component II"	
		/protein_id="BAB42563.1"	
		/db_xref="GI:13701269"	
		/translation="MAKLMNNEIKKVEORLEKAIKSDSVLPQASLHLLSSGKRV PAFVILSSQFCKDEQTSPTQVAVALELHMATLVHDDVIDKSDRRKRLISKWD"	
		99.0%; Score 1292.2; DB 1; Length 301550;	
		Best Local Similarity 99.4%; Pred. No. 1.3e-214;	
		Matches 1297; Conservative 0; Mismatches 8; Indels 0; Gaps 0;	
Qy	1	ATGACTAAACCTATAGTACCTATTGTAGGTAGGCTTAATGTAGCTAAATCTACAATTTT 60	
Db	13778	ATGACTAAACCTATAGTACCTATTGTAGGTAGGCTTAATGTAGCTAAATCTACAATTTT 13719	
Qy	61	AATAGAATAGTTGGAGACGTTTCGATTCTCAAGATACGCCAGGTGTACACCGAGAT 120	
Db	13718	AATAGAATAGTTGGAGACGTTTCGATTCTCAAGATACGCCAGGTGTACACCGAGAT 13659	
Qy	121	CGTATTATTTCTTCAGGTAGTGGTTAACACATGATTTCATATTATTATGACAGGTGT 180	
Db	13658	CGTATTATTTCTTCAGGTAGTGGTTAACACATGATTTCATATTATTATGACAGGTGT 13599	
Qy	181	ATTGAATTTGGTATGATCCACCATTCACAAACACAAATTAGAGCGGAGGAGAAATCGGCATA 240	
Db	13598	ATTGAATTTGGTATGATCCACCATTCACAAACACAAATTAGAGCGGAGGAGAAATCGGCATA 13539	
Qy	241	GATGAAGCGGATGTTATTATTATTTTATGTTAACTGCTGCTGAAGGATTGACACAAACCGAT 300	
Db	13538	GATGAAGCGGATGTTATTATTATTTTATGTTAACTGCTGCTGAAGGATTGACACAAACCGAT 13479	
Qy	301	GAATGGTGCCTCAAAATTTTATACAAATCTAAAAAACCGGTCGTATTAGCGGTTAACAAA 360	
Db	13478	GAATGGTGCCTCAAAATTTTATACAAATCTAAAAAACCGGTCGTATTAGCGGTTAACAAA 13419	
Qy	361	GTAGATAATATCGAATCGGTACAGAGTGTATGATTTCTATTATTAGGATTGGTGAA 420	
Db	13418	GTAGATAATATCGAATCGGTACAGAGTGTATGATTTCTATTATTAGGATTGGTGAA 13359	
		RESULT 2	
		AP003362/c	
		LOCUS	
		DEFINITION	
		Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete	
		sequence, section 5/9.	
		ACCESSION	
		AP003362 BA000017	
		VERSION	
		AP003362.2 GI:14247083	
		KEYWORDS	
		Staphylococcus aureus subsp. aureus Mu50	
		Staphylococcus aureus subsp. aureus Mu50	
		Bacteria; Firmicutes; Bacillales; Staphylococcus.	
		ORGANISM	
		Staphylococcus aureus subsp. aureus Mu50	
		Staphylococcus aureus subsp. aureus Mu50	
		Bacteria; Firmicutes; Bacillales; Staphylococcus.	
Qy	421	CCGTATCCAATATACAGGTCACATGCTTTAGTCTTGTGTACTGCTTGTAGATGCAGTGT 480	
Db	13358	CCGTATCCGATATACAGGTCACATGCTTTAGTCTTGTGTACTGCTTGTAGATGCAGTGT 13299	
Qy	481	TCTCATTTTGGTGAAGAGGAAGATCCTTATGATGAAGATACAAATTCGACTACTCAT 540	
Db	13298	TCTCATTTTGGTGAAGAGGAAGATCCTTATGATGAAGATACAAATTCGACTACTCAT 13239	
Qy	541	ATTGGAGCACCACCTAGTAAATCAAGTTTAGTAAATGCTATTTTAGTGAAGATCGC 600	
Db	13238	ATTGGAGCACCACCTAGTAAATCAAGTTTAGTAAATGCTATTTTAGTGAAGATCGC 13179	
Qy	601	GTTATCGTTTCTAATGTCAGGACACACGAGACGCTATTGATACAGAGTATAGTTAT 660	
Db	13178	GTTATCGTTTCTAATGTCAGGACACACGAGACGCTATTGATACAGAGTATAGTTAT 13119	
Qy	661	GATGGACAAGATTATGTTTAAATCGATCTGCTGCTATGCTGCTTAAAGAGGAAAGATATAT 720	
Db	13118	GATGGACAAGATTATGTTTAAATCGATCTGCTGCTATGCTGCTTAAAGAGGAAAGATATAT 13059	
Qy	721	GAATCAACTGAGAAATATTACGATTTAAGAGCTTTAAAGCGATTTGAACGTTCAAAATGTT 780	
Db	13058	GAATCAACTGAGAAATATTACGATTTAAGAGCTTTAAAGCGATTTGAACGTTCAAAATGTT 12999	
Qy	781	GTTTACTGCTTATTCATGACAAAGCATTTATTGAACAGATTAACGTTGTCAGGA 840	
Db	12998	GTTTACTGCTTATTCATGACAAAGCATTTATTGAACAGATTAACGTTGTCAGGA 12939	
Qy	841	TATGCACATCAACAGGTAAAGCAGTCGTGATGCTGCTGCTTAAATTAATGGGATCTGTCGAA 900	
Db	12938	TATGCACATCAACAGGTAAAGCAGTCGTGATGCTGCTGCTTAAATTAATGGGATCTGTCGAA 12879	
Qy	901	AAAGATAGTAAACGATGAAGAAATTTTGAAGATGAAGTACGTAAAGAAATTTTCAATTTT 960	
Db	12878	AAAGATAGTAAACGATGAAGAAATTTTGAAGATGAAGTACGTAAAGAAATTTTCAATTTT 12819	
Qy	961	GATTATGCACAAATTCCTTTGCTGCTGAAGAACCCACCAAGATACGTTACATTTATTC 1020	
Db	12818	GATTATGCACAAATTCCTTTGCTGCTGAAGAACCCACCAAGATACGTTACATTTATTC 12759	
Qy	1021	CCTTACATTAATGAACCAAGTGAACCAATATAAAACCGTGTTCAAAGTTCAACCTTTAAAT 1080	
Db	12758	CCTTACATTAATGAACCAAGTGAACCAATATAAAACCGTGTTCAAAGTTCAACCTTTAAAT 12699	
Qy	1081	GAAGTCGTTACTGATCAATTTCCATGACCCCTACACCAACAGAGTACAGCTTTG 1140	
Db	12698	GAAGTCGTTACTGATCAATTTCCATGACCCCTACACCAACAGAGTACAGCTTTG 12639	
Qy	1141	AATGCTCTTTTATGCAACACAGTTGCTATAGAACCCAGCATTTGTTGTTATTTGTTAAT 1200	
Db	12638	AATGCTCTTTTATGCAACACAGTTGCTATAGAACCCAGCATTTGTTGTTATTTGTTAAT 12579	
Qy	1201	GATGTAGAATTAATGCAATTTTCTTATTAACCGCTATTAGAGAATCAATCCGTCGCGCT 1260	
Db	12578	GATGTAGAATTAATGCAATTTTCTTATTAACCGCTATTAGAGAATCAATCCGTCGCGCT 12519	
Qy	1261	TTTGGTGTTCAGGTACACCAATTCATATTATAGCTCGAAAGAGA 1305	
Db	12518	TTTGGTGTTCAGGTACACCAATTCATATTATAGCTCGAAAGAGA 12474	

REFERENCE  
AUTHORS  
1 Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,  
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,  
Matsumaru,H., Maruyama,A., Murakami,H., Hosoyama,A.,  
Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C.,  
Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Tabuzaki,J.,  
Kanehisa,N., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,  
Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramatsu,K.  
Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*  
Lancet 357 (9264), 1225-1240 (2001)  
21311952  
11418146  
REFERENCE  
PUBMED  
2 (bases 1 to 346900)  
Ohta,T.  
Direct Submission  
Submitted (28-FEB-2003) Toshiko Ohta, University of Tsukuba College  
of Medical Technology and Nursing, Department of Medical  
Technology; 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan  
(E-mail:tohta@esakura.cc.tsukuba.ac.jp, Tel:81-298-53-3454,  
Fax:81-298-53-3454)  
On May 29, 2001 this sequence version replaced gi:13875626.  
COMMENT  
FEATURES  
source  
1. 346900  
/organism="Staphylococcus aureus subsp. aureus Mu50"  
/mol\_type="genomic DNA"  
/strain="Mu50"  
/sub\_species="aureus Mu50"  
/db\_xref="taxon:158878"  
12. 263  
/gene="SAV1313"  
12. 263  
/gene="SAV1313"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA857475.1"  
/db\_xref="GI:14247084"  
/translation="MUTLKDEQVIVYIYEPEDDISNGKSVTFNKDAEVIDFS  
LSEIENEYPMLYRNKSFVVRDFIEKQFPENYKIAY"  
987. 1181  
/gene="SAV1314"  
987. 1181  
/gene="SAV1314"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA857476.1"  
/db\_xref="GI:14247085"  
/translation="MNLQSYIGKLVKLTNNKILIGKVIDFDKQVDFGYSNIEID  
TGRITVDISENKIKTIVLQK"  
complement(1725..2750)  
/gene="SAV1315"  
complement(1725..2750)  
/gene="SAV1315"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BA857477.1"  
/db\_xref="GI:14247086"  
/translation="MISFENDYLEGAHEKVKRLVDTNLVQASGYGDFQTAQAIKI  
KDTIDPNATIRFLVGCTQTNQVIVNSMLESYEGVISADTGHVAVHGGALIEYSGHKV  
ITIPSGKGVASDVTYMETFKSDPKKDHVFGMVYISHPTVEYGTLYSKSELEILC  
KVKQOLPULPMGARGLGLMSDOSDMTIDIAKYCDVFIYIGCTKIGALCGEALVPT  
KNEPQFTRIKHICALLAKGLGLOFLELPTDNLVFNISRHVAIEAMKMKDGFKN  
KGYRLVDSPTNQOFTILSNEIALEQKVFAVWEKYDDQHRVVRVATSWATTEENL  
NKLELI"  
3020. 3217  
/gene="SAV1316"  
3020. 3217  
/gene="SAV1316"  
/codon\_start=1  
/transl\_table=11  
gene  
CDS  
/product="hypothetical protein"  
/db\_xref="GI:14247087"  
/translation="MEQIKUKTFTAEITLLEKNINAPLSSBEATNLKLVNITIKEIE  
EXTFNNSEEPNAILTUSYNK"  
3238. 4755  
/gene="SAV1317"  
3238. 4755  
/gene="SAV1317"  
/codon\_start=1  
/transl\_table=11  
/product="cardiolipin synthetase homolog"  
/protein\_id="BA857479.1"  
/db\_xref="GI:14247088"  
/translation="HFFVPPFIGEKTHRYTFSNDLGLTPTIILAIGFINLVLAFLII  
FLERNRRTASSTWALFVFLVPLIGLFLYFGRTVSARKLNKNGNVLTFDGLLK  
QOIESDFDNGTGNQVQKHLDLVSMLLMDQDFLTENKVDHFIDGNLDYDVLKD  
IKNAKEYIHLEYTTFALDGLGKRILHALEKIVKQGLEVKILYDVGSKNVKNANPDHF  
KSLAGEVAFASKLPLNFRMNNRHKIIVIDQGLYVGGFNGIDGVEYGLGRLGYV  
RDTLRIQGDVADALQRLRFLDWNNSQHRPQFEDYKVPKNGPLGNSPIQIAAGGP  
ASDWHOIEGYVTKIMSAAKSVLYQSPYFIPDNYSYINAIKIAKSCVDVHLMIPCKPD  
HPLVTMATESNADLLSSGVKIITYENGFIHSMCLIDDEIVSVGTANMDFSEFLNP  
EVNAFVDENLAKDLRVAYEHDTKSKQLTKESYANRPLSVAFKESLAKLVSPIL"  
4939. 5436  
/gene="SAV1318"  
4939. 5436  
/gene="SAV1318"  
/note="ATP-binding protein"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter homolog"  
/protein\_id="BA857480.1"  
/db\_xref="GI:14247089"  
/translation="MIOISNINKSENKRCVLRKNISDFIOGKCIALICNGAGKSTLI  
DILIGNWNANSGEIFDKKLQSENKSIPOKTMFPDOLKVIIEINLVQSPYENPLPL  
BEIIEILKFDSSQLNQFVNLGGQORLLDFVLISLGGQLILDEPTSTMDIEIREY  
FWSII"  
5461. 5838  
/gene="SAV1319"  
5461. 5838  
/gene="SAV1319"  
/note="ATP-binding protein"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter homolog"  
/protein\_id="BA857481.1"  
/db\_xref="GI:14247090"  
/translation="MLYTSYIEVERMSDKIILIEGEIILNDSTSHRTNQSQIT  
LSDEYIRKLKLDKDLVQIKNHNGTIKIITSNNVDITLYLOQLHINLDDIEIQKYSIV  
DSYFNKKQKGSNYDTKLENRI"  
5807. 6538  
/gene="SAV1320"  
5807. 6538  
/gene="SAV1320"  
/note="similar to ABC transporter integral membrane  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BA857482.1"  
/db\_xref="GI:14247091"  
/translation="MILSYLKIIEFKVIMRKTKTLILSPVIFVILFTSILEPEDV  
KPKYKVMYSMTVSYLSLSPSLTFLDIIINEKQMRQRLVPTPTFTSYISKVVK  
TMLQFATILVFMVGHFYKGVASAVHLESGIFLWLGASLIITFGILPULLNDIQK  
TSALANVLTGLAVLGGFPIINTFPNNLQHVANVLPSTHLKRLGVADIASNNHNLIS  
FAILLVLTGSIIAVYICISFKRAE"  
6542. 7633  
/gene="SAV1321"  
6542. 7633  
/gene="SAV1321"  
/note="similar to two-component sensor histidine kinase"  
/codon\_start=1  
/transl\_table=11  
gene  
CDS  
/product="hypothetical protein"  
/db\_xref="GI:14247092"  
/translation="MILSYLKIIEFKVIMRKTKTLILSPVIFVILFTSILEPEDV  
KPKYKVMYSMTVSYLSLSPSLTFLDIIINEKQMRQRLVPTPTFTSYISKVVK  
TMLQFATILVFMVGHFYKGVASAVHLESGIFLWLGASLIITFGILPULLNDIQK  
TSALANVLTGLAVLGGFPIINTFPNNLQHVANVLPSTHLKRLGVADIASNNHNLIS  
FAILLVLTGSIIAVYICISFKRAE"  
6542. 7633  
/gene="SAV1321"  
6542. 7633  
/gene="SAV1321"  
/note="similar to two-component sensor histidine kinase"  
/codon\_start=1  
/transl\_table=11



SOURCE	Staphylococcus aureus	
ORGANISM	Staphylococcus aureus	
REFERENCE	Bacteria; Firmicutes; Bacillales; Staphylococcus.	
AUTHORS	1 Masignani,V.C., Mora,M.C. and Scarselli,M.C.	
TITLE	Staphylococcus aureus proteins and nucleic acids	
JOURNAL	Patent: WO 02094868-A 5631 28-NOV-2002;	
	Chiron spa (IT)	
FEATURES	Location/Qualifiers	
source	1. 1308	
	/organism="Staphylococcus aureus"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:1280"	
BASE COUNT	450 a 184 c 278 g 396 t	
ORIGIN		
Query Match	98.9%; Score 1290.6; DB 6; Length 1308;	
Best Local Similarity	99.3%; Pred. No. 9.9e-214;	
Matches 1296;	Conservative 0; Mismatches 9; Indels 0; Gaps 0;	
QY	1 ATGACTAAACCTATAGTAGCTATTGTAGGTAGCCCTAATGTAGTAAATCTACAATTTT 60	
Db	1 ATGACTAAACCTATAGTAGCTATTGTAGGTAGCCCTAATGTAGTAAATCTACAATTTT 60	
QY	61 AATAGATAGTTCGAGAACGTTTTCGATTGTGAAGATACGCCAGGTGTAAACAGAGAT 120	
Db	61 AATAGATAGTTCGAGAACGTTTTCGATTGTGAAGATACGCCAGGTGTAAACAGAGAT 120	
QY	121 CGTATTTATCTTCAGGTGAGTGGTTAAACACATGATTTCATATATTATTGATACAGTGGT 180	
Db	121 CGTATTTATCTTCAGGTGAGTGGTTAAACACATGATTTCATATATTATTGATACAGTGGT 180	
QY	181 ATTGAATTTGTTGATGACCACTTCCAAACACAATTTAGAGCCGACGAGATCCGCATA 240	
Db	181 ATTGAATTTGTTGATGACCACTTCCAAACACAATTTAGAGCCGACGAGATCCGCATA 240	
QY	241 GATGAGCGGATGTTATTTATTTATGTTTAAAGTCCGCGTGAAGGATTGACACAAAGCGAT 300	
Db	241 GATGAGCGGATGTTATTTATTTATGTTTAAAGTCCGCGTGAAGGATTGACACAAAGCGAT 300	
QY	301 GAAATGGTGGTCAAAATTTTATACAAATCTAAAAAACCGGTGCTATTAGCGGTTAAACAAA 360	
Db	301 GAAATGGTGGTCAAAATTTTATACAAATCTAAAAAACCGGTGCTATTAGCGGTTAAACAAA 360	
QY	361 GTAGATATATGGAATGCGTACAGAGCGTGTATGTTCTTATTCATTCATTAGGATTGGTGAA 420	
Db	361 GTAGATATATGGAATGCGTACAGAGCGTGTATGTTCTTATTCATTCATTAGGATTGGTGAA 420	
QY	421 CCGTATCAATATCAGGGTCACATGGTTTAGGTCTTTGGTGACTTGTATAGATGCAGTTGTT 480	
Db	421 CCGTATCCGATATCAGGGTCACATGGTTTAGGTCTTTGGTGACTTGTATAGATGCAGTTGTT 480	
QY	481 TCTCATTTTGGTGAAGAGGAAGATCCCTTATGATGAAGATACAAATTCGACTATCCATT 540	
Db	481 TCTCATTTTGGTGAAGAGGAAGATCCCTTATGATGAAGATACAAATTCGACTATCCATT 540	
QY	541 ATTGAGCAGCAACCGTAGCTTAATCAAGTTTATGTAATGCTATTTTAGGTGAAGATCGC 600	
Db	541 ATTGAGCAGCAACCGTAGCTTAATCAAGTTTATGTAATGCTATTTTAGGTGAAGATCGC 600	
QY	601 GTTATCGTTTCTAATGTTGAGGAGCAACGAGAGCGCTATTATGATACAGAGTATAGTTAT 660	
Db	601 GTTATCGTTTCTAATGTTGAGGAGCAACGAGAGCGCTATTATGATACAGAGTATAGTTAT 660	
QY	661 GATGACAAAGATATTGTTTAAATCGATCTGCTGGTATCGGTAAAAAGGAAAGATATAT 720	
Db	661 GATGACAAAGATATTGTTTAAATCGATCTGCTGGTATCGGTAAAAAGGAAAGATATAT 720	
QY	721 GAATCAACTCAGAAATATTCAAGTATTAAAGAGCTTTAAAAAGCGATTGAACGTTCAAAATGTT 780	
Db	721 GAATCAACTCAGAAATATTCAAGTATTAAAGAGCTTTAAAAAGCGATTGAACGTTCAAAATGTT 780	
QY	781 GTTTTACTGCTATTGATGACAGAACAGCATTTATTGAACAGATAAACGTTGTCAGGA 840	
RESULT 4		
AP004827/c	333750 bp DNA linear BCT 24-APR-2003	
LOCUS	Staphylococcus aureus subsp. aureus MW2 DNA, complete genome,	
DEFINITION	strain:MW2, section 6/10.	
ACCESSION	AP004827 BA000033	
VERSION	AP004827.1 GI:21204509	
KEYWORDS		
SOURCE	Staphylococcus aureus subsp. aureus MW2	
ORGANISM	Staphylococcus aureus subsp. aureus MW2	
REFERENCE	Bacteria; Firmicutes; Bacillales; Staphylococcus.	
AUTHORS	1 Baba,T., Takeuchi,F., Kuroda,M., Yuzawa,H., Aoki,K., Oguchi,A., Nagai,Y., Iwama,N., Asano,K., Nalim,T., Kuroda,H., Cui,L., Yamamoto,K. and Hiramatsu,K.	
TITLE	Genome and virulence determinants of high virulence community-acquired MRSA	
JOURNAL	Lancet 359 (9320), 1819-1827 (2002)	
MEDLINE	22040717	
PUBMED	12044378	
REFERENCE	2 (bases 1 to 333750)	
AUTHORS	Director-General, Biotechnology Center, Aoki,K., Oguchi,A., Nagai,Y., Asano,K., Iwama,N., Baba,T., Kuroda,M., Hiramatsu,K. and Kikuchi,H.	
TITLE	Direct Submission	
JOURNAL	Submitted (06-MAR-2002) Director-General, Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center, 2chome 49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail:bioelite.go.jp, URL:http://www.bio.nite.go.jp/, Tel:81-3-3481-1933, Fax:81-3-3481-8424)	
FEATURES	Location/Qualifiers	
source	1. 333750	
	/organism="Staphylococcus aureus subsp. aureus MW2"	
	/mol_type="genomic DNA"	
	/strain="MW2"	
	/db_xref="taxon:196620"	
	complement(278..619)	
gene		

```

/ gene="MW1341"
/ complement(278..619)
/ gene="MW1341"
/ note="ORFID:MW1341"
/ codon_start=1
/ trans1_table=11
/ product="hypothetical protein"
/ protein_id="BAB95206.1"
/ db_xref="GI:21204510"
/ translation="MTNITKEVFDNLEQIDVPAKNKTLGSSEAKPYLDVHSHIIDY
PKOVNITGNIDPKLNQIPVVPVMEKERYDVMIERKHYFMGYRQMKTFTKELIKMNA
SYTRLNKQV"
/ complement(624..1283)
/ gene="nth"
/ complement(624..1283)
/ gene="nth"
/ note="ORFID:MW1342"
/ codon_start=1
/ trans1_table=11
/ product="endonuclease-like protein"
/ protein_id="BAB95207.1"
/ db_xref="GI:21204511"
/ translation="MYSKAKALEMIDVIANMFPDACELKNDNFELTIYAVLSAQT
DVLNVATTELEFKKYPEDYLAVSDEELMNDIRSLGRNKAKNIRKLAQSLIDQFN
GEIPTQKLESLAGVGRNTANVMSVAFDEPSLAVDTHVERVSKRIGINRWKDNVRQ
VEDRLCSVIPDRNRSHQLIFFGRYHCLARKPKDCICPLEDCREGQRKYRASLKE
A"
/ complement(1273..1959)
/ gene="MW1343"
/ complement(1273..1959)
/ gene="MW1343"
/ note="ORFID:MW1343"
/ product="hypothetical protein, similar to chromosome replication
initiation protein dnaB"
/ codon_start=1
/ trans1_table=11
/ protein_id="BAB95208.1"
/ db_xref="GI:21204512"
/ translation="MDKYOLKARPVVIRRELDHYSDGLDQDLVLLKLIYASETS
NKQPSITELQKSTMPDRITNVIONLQRELLQVOKDEGRFTYMWLDPDFEKL
SHILKQSMETKEONSKFKVLFEDTAPRLSPYEIETLQWIDVKNHSAIIQ
AALDEANSKLKLSFKYMDRILLMWKNVYIDDSRKIRERFNAPMTHTVTKVTPFD
WLNGLNDGK"
/ complement(1287..3579)
/ gene="asn"
/ complement(1287..3579)
/ gene="asn"
/ note="ORFID:MW1344"
/ codon_start=1
/ trans1_table=11
/ product="asparaginyl-tRNA synthetase"
/ protein_id="BAB95209.1"
/ db_xref="GI:21204513"
/ translation="MKTITQAKDHLNDQVITGAWLTNRSSGKIAFLQLRDGTGPMQ
GVVXSEDEEVFKLAKETIOESSLYVTGTITEDNRSDLYGMQVKSIEVISAHDYP
ITPKNGIEFLMDHRLHRLSKQHOAVMKIRNEVIRATYFENKDGFTKVDPTLAS
APEGTSELEFRTYFQDAFLSQSGQLYLEAAMAAGKVFSPGPTFRAEKSSTRHLIE
FMIEGMAFTNHAESLEIQEYVTHVVSYLENCKLEKILERTDTSKLEKVPATPPR
ISVDDAIEFLKSGEDDIEMGEDFCAPHETAIANHYDLVPVITNYPTKIKPFYKQPNP
ENETVLCADLIAPEGYELIGGSEFVDDLELLEQVRKEHGLDEAYSYVLDLRRYGS
VPHCGGLGLETVAVISGVENHRETAPFPLLNRLYP"
/ complement(1901..6594)
/ gene="ding"
/ complement(1901..6594)
/ gene="ding"
/ note="ORFID:MW1345"
/ codon_start=1
/ trans1_table=11
/ product="probable ATP-dependent DNA helicase ding"
/ protein_id="BAB95210.1"
/ db_xref="GI:21204514"
/ translation="MGMTYAVVLETTGNGQLDFDDIIQIGITFVRNQNIIQTHYSKI
RTWLEIPPIQALTSTERNLQQAIFYNQVAQEYIDKIDCIPVAHNVDFDLNFIKA
FKDCNIOYRPKKVITDLEIFKIAFPDTSKYSQVLSAEAHGIIITLANHRADEADAATTAK
LMILAFEFKPLDITLKQYIYSKQYLDLYDIFFEMVROYDAKPLDKSVYKFEQII
YRKQVDFKPTNTYNGSLKSYKAVDQGLCYRPOOLYLAETIILDQLHSEKAMIEA
SLGSGKSLAYLLAALMYNIETGRHVMISTNTKLSQOLLEKDI PANNEALNFKINALL
IKSKSDYISLGLISOILKDDTSNVEYNILKMOLLWIITETPSGDIQELNKGQKMYF
DQKTEYVPAHDVHYNFYIKRANOIQTGITHNAHLIHSVDENSIYOLPDDCLVDEA
HRLPYALNOVTNELSVADIKYOLGIGKNEKELUKAIDOLEKORILEKLDIAPIDI
ECLKASNEHELNEOLFSTIFTIINDSDVDDDIHRFNWTFPTKOLIKOLHAIID
KLNKTEIFNCISHKTVKSKOLLKDKFNIBOSLAAGHTSFISIKNSQKSTIR
LYADIYAVDLTKOVLKFKSLIFISGTLKFNHSEAFKQFNKDVHFNFTFEVNTSL
QSAKTSVPSVDASVQYKIDVEYASIVSYIETIITSSKLVLTFTSKMMHMQ
DKMLHPEFEDYVVLTOQONKTKIQQFNFDKAILLGTSTFFGFGFQANGKCVM
IAKLPMHKNHAKYMLMDEFTSTFKEYVLPAVTRFQGLRLIRSENDRGIIIVSPD
DRLINSYKNFFEQTLNRYQRKQKDIQQFCKLLRQIQKKK"
/ gene="MW1346"
/ complement(6618..7589)
/ gene="MW1346"
/ note="ORFID:MW1346"
/ codon_start=1
/ trans1_table=11
/ product="truncated hypothetical protein"
/ protein_id="BAB95211.1"
/ db_xref="GI:21204515"
/ translation="MSKYSQVLOLLYKKNPNYISGQSTAESLNTSRTAVKVVIOLEK
LEGCKIDSVNHKGHLLQPLDIWQGIIDQYTKSSALDFSEVDSIDSTOLAKKSL
VGNOSFFILLSDQTKGRGFENHSSSGCGLMHSVYLRNPAFMSHLSKFNALIG
IRDAIQHFSQDEVKYKPNDIYIDNGKVGFLTEYANNNDGEATIGIGINCLOOLE
NFDSEIRHRTSVQLHDKNLRQTFLERLQJEIEKRYNQFLPFPFSIREEYTAASN
IWNRTLLFTENDQFKQAIDLDYDGYLIVRDEAGESHRLISADIOF"
/ complement(7576..8778)
/ gene="MW1347"
/ note="ORFID:MW1347"
/ product="hypothetical protein, similar to poly(A) polymerase"
/ codon_start=1
/ trans1_table=11
/ protein_id="BAB95212.1"
/ db_xref="GI:21204516"
/ translation="MKSLSFEQARPILEDIQNGFEAYVVGSVRYVGMRIHIDI
TTSATPATESISFHTIPVKGEGTINNVFNDENYEVITFRAEEDYVDRHSPDGVTY
RDLYEDLQRDEATMNAIMDTAYKLYDFDQGDINNRRIIRTVTAEERFPODGLRM
ICLRFQSLQSDFIATETFAEMRIQMAIDINFLSIRIELTKMRGINVKESFNHLKS
LKAFNMPYFELHDMQINVTFAIDLELLIAVSVKFOINYSKPLKLSNRQVADINO
YIQIMNALPSIITKQOLKMPYDYDTHLIKNMVAADVIKANDIQGHPLVNLQTIID
ETLHRLPMHNRKDMWNGVGLMAHLNAKSGPLKDVLRQIEIAIVTGKVSNEETEILK
WVDNHVKI"
/ complement(8783..9925)
/ gene="MW1348"
/ complement(8783..9925)
/ gene="MW1348"
/ note="ORFID:MW1348"
/ product="hypothetical protein, similar to lipopolysaccharide
biosynthesis-related pr homolog"
/ codon_start=1
/ trans1_table=11
/ protein_id="BAB95213.1"
/ db_xref="GI:21204517"
/ translation="MKIGITCPSMGCGSIIATELGIKLAERGHEVHFTSNIIPFIR
KPLPMHIFHVEVNVQAVFOYPPYDITLSTKIAEVIKGYDOLLHMYAVPHATCGIL
ARENSGDKIKIMTTLHGDTITVLGYDHSIQGAIKRGIKSDIVTSVSKSLAETHETII
ETNKEIIPYFVRENEFTKHTALKSQFIAPDEKVLHVSFNPQVVKRITDIETFF
AKVREKIPSKILLGDGPBLVPMRQLTKELNVEEDVLGKQCVSFYQLDLVLL
SEKESFGLTLEAMKTGPIGSGNAGGKEVIKHGCTGTGVVDGDCSDASDAIRLLE
DKVLYNKLOKNMLADIAERFSGELITDQVEYYVYOKMLNEHKNKSGE"
/ complement(10167..10484)
/ gene="MW1349"
/ complement(10167..10484)
/ note="ORFID:MW1349"
/ codon_start=1
/ trans1_table=11

```

/product="conserved hypothetical protein"	
/db_xref="GI:21204518"	
/translation="MSKVKQREVEYIGORFQVSPFLANLRLTEVCELAREIN	
HTYGEKKKKSEERANTKAEIGDNLFLVLLCLANSMDINTESFNEMKFNTRDKRF	
ERK	
gene	Complement(10820..11518)
CDS	/gene="MW1350"
	complement(10820..11518)
	/gene="MW1350"
	/note="ORFID:MW1350"
	/codon_start=1
	/transl_table=11
/product="conserved hypothetical protein"	
/protein_id="BAB95215.1"	
/db_xref="GI:21204519"	
/translation="MSSIFTHIYFVILMWPLVLAHQHKVSKYKYSQVRSKTKGR	
EVALLEIHANGIYDEVYKGEFLDHYDPKKVLSLSPANTYDRPSVAGTAIAAHEVG	
HAIQDHQYFLFRALVPAVLGSSLSYHIMLGIILTAIGSAGSTALWIGALM	
Query Match	
Best Local Similarity 99.3%; Pred. No. 2.3e-214;	
Matches 1296; Conservative 0; Mismatches 9; Indels 0; Gaps 0;	
Qy	1 ATGACTAACTATAGTACGCTATGTTAGGTAGCGCTAATCTAGGTAAATCTCAATTTT 60
Db	
Qy	24351 ATGACTAACTATAGTACGCTATGTTAGGTAGCGCTAATCTAGGTAAATCTCAATTTT 24292
Db	
Qy	61 AATAGATAGTTGGAGACGGTGTTCGATTTGGAAGATACGCCAGGTCTAACACGAGAT 120
Db	
Qy	24291 AATAGATAGTTGGAGACGGTGTTCGATTTGGAAGATACGCCAGGTCTAACACGAGAT 24232
Db	
Qy	121 CGTATTATCTTCACGTGACGTGGTTAACACATGATTTCAATATATTATCATACAGTGT 180
Db	
Qy	24231 CGTATTATCTTCACGTGACGTGGTTAACACATGATTTCAATATATTATCATACAGTGT 24172
Db	
Qy	181 ATTGAATTTGGTATGACGACATTCACAAACACAAATAGACGCGACGACAAATCCCAT 240
Db	
Qy	24171 ATTGAATTTGGTATGACGACATTCACAAACACAAATAGACGCGACGACAAATCCCAT 24112
Db	
Qy	241 GATGACGCGATGTTATTATTTTATGTTTAACTGCGTGAAGGATTGACACAAACGAT 300
Db	
Qy	24111 GATGACGCGATGTTATTATTTTATGTTTAACTGCGTGAAGGATTGACACAAACGAT 24052
Db	
Qy	301 GAAATGGTGGCTCAAAATTTATACAAATCTAAAAACCGTGCCTATTACCGGTTACAAA 360
Db	
Qy	24051 GAAATGGTGGCTCAAAATTTATACAAATCTAAAAACCGTGCCTATTACCGGTTACAAA 23992
Db	
Qy	361 GTAGATAATATGGAATCGGTACAGACGTGTATGATTTCTATTCTATTAGGATTGGTGAA 420
Db	
Qy	23991 GTAGATAATATGGAATCGGTACAGACGTGTATGATTTCTATTCTATTAGGATTGGTGAA 23932
Db	
Qy	421 CCGTATCCAATATCAGGTCACATGGTTTGGTCTTGGTACCTTGTAGATGACGTTGT 480
Db	
Qy	23931 CCGTATCCAATATCAGGTCACATGGTTTGGTCTTGGTACCTTGTAGATGACGTTGT 23872
Db	
Qy	481 TCTCATTTTGGTGAAGGAGAGATCCCTTATGATGAAGATACAAATTCGACTATCCATT 540
Db	
Qy	23871 TCTCATTTTGGTGAAGGAGAGATCCCTTATGATGAAGATACAAATTCGACTATCCATT 23812
Db	
Qy	541 ATTGACGACCAACGATAGGTAAATCAAGTTTATGATTAATGCTATTTTAGGTGAAGATCGC 600
Db	
Qy	23811 ATTGACGACCAACGATAGGTAAATCAAGTTTATGATTAATGCTATTTTAGGTGAAGATCGC 23752
Db	
Qy	601 GTTATCGTTTCTTAATGTTGACGAGCAACGAGACGCTATTTCATACAGATATAGTTAT 660
Db	
Qy	23751 GTTATCGTTTCTTAATGTTGACGAGCAACGAGACGCTATTTCATACAGATATAGTTAT 23692
Db	
Qy	661 GATGACAGATATATGTTTTTAAATCGACTGCTGCTGATGCGTAAAAAGGAAAGTATAT 720
Db	
Qy	23691 GATGACAGATATATGTTTTTAAATCGACTGCTGCTGATGCGTAAAAAGGAAAGTATAT 23632
Db	
Qy	721 GAATCACTGAGAAATATTCAGTATTAAAGACCTTTAAAGACGATTGAACGTTCAATGTT 780
Db	

Db	23631	GAATCACTGAGAAATATTCAGTATTAAAGACCTTTAAAGACGATTGAACGTTCAATGTT	23572
Qy	781	GTTTTAGTGGTTATTGATGCGAGAACAGGCAATTATTGAACAGATAAACGCTGTTGCAGGA	840
Db	23571	GTTTTAGTGGTCTATAGATGCGAGAACAGGCAATCATTGAACAGATAAACGCTGTTGCAGGA	23512
Qy	841	TATGCACATGAACAAGGTAAAGCAGTCGTGATTCGTCTAAATAAATAAATGGGATCTCTGGNA	900
Db	23511	TATGCACATGAACAAGGTAAAGCAGTCGTGATTCGTCTAAATAAATAAATGGGATCTCTGGNA	23452
Qy	901	AAAGATAGTAAACGATGAAGAAATTTGAAGATGAAGTACGTAAAGAAATTTCAATTTTFA	960
Db	23451	AAAGATAGTAAACGATGAAGAAATTTGAAGATGAAGTACGTAAAGAAATTTCAATTTTFA	23392
Qy	961	GATTATGCACAAATTCCTTTCTCTCTAAAGACGACAGATTTAGTACATATTTC	1020
Db	23391	GATTATGCACAAATTCCTTTCTCTCTAAAGACGACAGATTTAGTACATATTTC	23332
Qy	1021	CCTTACATTAATGAAGCAAGTGAACCAATTAACCAAGCTGTTCAAAAGTTCACATTTAAAT	1080
Db	23331	CCTTACATTAATGAAGCAAGTGAACCAATTAACCAAGCTGTTCAAAAGTTCACATTTAAAT	23272
Qy	1081	GAAGTCGTTTACTGATGCAATTTCCATGAACCCCTACACCAACAGGACAGCTTTG	1140
Db	23271	GAAGTCGTTTACTGATGCAATTTCCATGAACCCCTACACCAACAGGACAGCTTTG	23212
Qy	1141	AATGTCCTTTTATGCAACACAGTTGCTATAGAACCCAGACATTTGTTGTTATTAT	1200
Db	23211	AATGTCCTTTTATGCAACACAGTTGCTATAGAACCCAGACATTTGTTGTTATTAT	23152
Qy	1201	GATGATAGATTAATGCAATTTTCTTATAAACGCTATTATAGAGATCAAAATCCGTCGCT	1260
Db	23151	GATGATAGATTAATGCAATTTTCTTATAAACGCTATTATAGAGATCAAAATCCGTCGCT	23092
Qy	1261	TTTGGTTTGAAGGTACACCAATTCATATTATAGCTCGAAGAGA	1305
Db	23091	TTTGGTTTGAAGGTACACCAATTCATATTATAGCTCGAAGAGA	23047

RESULT 5	AF270032/c	3269 bp	DNA	linear	BCT 01-AUG-2000
LOCUS	Staphylococcus epidermidis strain SRI clone step.1042f07 genomic				
DEFINITION	sequence.				
ACCESSION	AF270032				
VERSION	AF270032.1	GI:9623936			
KEYWORDS	Staphylococcus epidermidis				
SOURCE	Staphylococcus epidermidis				
ORGANISM	Bacteria; Firmicutes; Bacillales; Staphylococcus.				
REFERENCE	1 (bases 1 to 3269)				
AUTHORS	Kimberly, W.J., Taylor, J. David., Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelsen, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuler, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.				
TITLE	Transposon-mediated sequencing of the Staphylococcus epidermidis				
JOURNAL	genome				
REFERENCE	2 (bases 1 to 3269)				
AUTHORS	Taylor, J. David., Kimberly, W.J., Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelsen, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuler, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.				
TITLE	Direct Submission				
JOURNAL	Submitted (22-MAY-2000) Departments of Genomic Sciences and				
FEATURES	Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore				
source	Drive, Research Triangle Park, North Carolina 27709-3398, USA				
	1. 3269				
	/organism="Staphylococcus epidermidis"				



```
/mol_type="genomic DNA"
/strain="SR1"
/db_xref="taxon:1282"
/cloned_ref="step.1042107"
BASE COUNT      986 a   616 c   451 g   1216 t
ORIGIN
Query Match      70.5%; Score 919.4; DB 1; Length 3269;
Best Local Similarity 81.5%; Pred. No. 1.3e-149;
Matches 1064; Conservative 0; Mismatches 241; Indels 0; Gaps 0;
OY      1  ATGACTAAACCTATAGTACTAGTATTGTAGTGGCTTAATGTAGGTAAATCTACAAATTTT 60
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2804  ATGACTAAACCGATAGCAATTTGAGGAAACCAATAGGAGAAATCTACAAATTTT 2745
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      61  AATGAAATAGTTCGGAACGCTGTTTCGATTTGTTGAAGATACGCCAGCGTGAACACGAGAT 120
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2744  AATGAAATAGTTCGCGCAAGCTATATCTATCGTAGAAGATACGCCAGCGCTTACTCGTAT 2685
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      121 CGTATTTATCTTCAGTGAGTGGTTAAACACATGATTTCAATATATTATGATACAGGTGT 180
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2684  CGTATTTATCTTCAGTGAGTGGTTAACTCATGATTTAACTCATGATTTAACTCATGATTTAACT 2625
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      181 ATTGAAATGGTGATCGACCATCTCCAAACACAAATTTAGACGCCAGCGCAAAATCGCCATA 240
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2624  ATTGAAATGGTGATCGCTCTCTTTTCAAACTCAAAATCCCGTCACAGCGCAAAATAGCAATT 2565
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      241 GATGAAGCGGATGTTATTTATGTTTATGTTTAACTGCGTGAAGGATTTGACACAAAGCGAT 300
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2564  GAAGACGATGTCATCTATTTTATGTTCAATTTAGAGAAAGGACTTACACAAAGTGAC 2505
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      301 GAAATGTCGCTCAAAATTTATACAAATCTAAACACCGTCTGATTTAGCGGTTAAACAAA 360
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2504  GAAATGTCGCAAAATGCTTTATAAATCTAAGAACTGTTGTTAATTTAGCTGTGAATANA 2445
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      361 GTAGATAATATGGAATGGCTACAGACGTGTATGATTTCTATTCTATTAGGATTTGGTGAA 420
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2444  GTTGATAATCTGGAATGGCTAATGATATCTATGATTTCTATTCTTTTAGGCTTTGGAGAT 2385
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      421 CCGTATCCAATATCAGGTCACATGTTTGTAGTCTTGTGCTTGTAGCTTGTAGCTGCAATGTT 480
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2384  CCATATCCTATTTCTGTTCCATCGATTTAGCACTTGGAGATTTGCTAGATGCAATGTT 2325
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      481 TCTCATTTTGGTGAAGAGAGAGATGCTTTATGATGAAGATACAAATTCGACTATCCATT 540
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2324  GAAACCTTTAATAAGAAATCAGAAGATCCTTTATGACGAAGATACGATACGCTTTCTATC 2265
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      541 ATTGACGACCAACCTAGTAAATCAAGTTTATGTAATGCTATTTTATGTTGAAGATATAT 600
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2264  ATCGGTAGACCTAATGTTGTTAACTAGCTTGGTCAATGCTATTTTAGGCGAAGACGT 2205
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      601 GTTATGCTTTCTAATGTTTCAGGACACGACGAGCGTATTTGATACAGATATAGTTAT 660
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2204  GTTATGCTTTCTAATGTTTCAGGACACGAGTATTTGATACAGATATAGTTATAGCTTCTAT 2145
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      661 GATGACGACGATTTATGTTTAACTCGATCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTG 720
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2144  GATGACGACGATTTATGTTTAACTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAT 2085
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      721 GAATCAACTCAGAAATATTACGATTTAAACGCTTTAAAGCGATTTGAACGTTTCAATGTT 780
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2084  GAATCAACTCAGAAATATTCTGTTTAACTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2025
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      781 GTTTTAGTGGTTATTGATGACGACGACGATTTTGAACAGATTAACAGTGTGTCAGGA 840
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2024  GTATTAGTGGTTATTGATGACGACGATTTTGAACAGATTAACAGTGTGTCAGGA 900
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      841 TATGACATCAACCAAGGTAAAGCAGTCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1964  TATGACATCAACCAAGGTAAAGCAGTCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1905
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY      901 AAACATAGTAAACGATGAAGAAATTTGAAGATGAAGTACGTTAAAGATTTCCAAATTTT 960
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```







Db	297004	ATTGAAATTTGGAGATGCTCTCTTTTCAAACTCMAATCGTGCCACAGCAGAGAAATAGCAATTT	296945
QY	241	GATGAAGCGGATGTTATTTATTTTATGTTTAACTGCTGCTGAAAGGATTGACACAAAGCGAT	300
Db	296944	GAAGAAGCAGATGTCATCTATTTTATGTTCAATGTTATGAGAAAGGACTTACACAAAGTGAC	296885
QY	301	GAATGGTGCCTCAAAATTTTATACAAATCTAAAAAACCGGTCGTATTAGCGGTTTAAACAA	360
Db	296884	GAATGGTGCACAAATCTTTATAAATCTAAGAAACCTCTGTTATTAGCTGTGAATAAA	296825
QY	361	GTAGATAAATGSGAAATCGGTACAGAGGTGTGATTTCTTATTCATTAGATTTTGGTGAA	420
Db	296824	GTTGATATCTTGAATTCGGTATGATATCTATGATTTCTTTAGCCTTTGGAGAT	296765
QY	421	CCGTATCCAAATCAGGGTGCACATGGTTTAGTCTTGGTGACCTTTGTTAGATGCAAGTTGTT	480
Db	296764	CCATATCCTATTTCTTGGTTTCACATGGGATTAGGACTTTGGAGATTTGCTAGATGCAGTTGTT	296705
QY	481	TCTCATTTTGGTGAAGAGAGAAAGATCCTTTATGATGAAGATACAAATTCGACTATCCATT	540
Db	296704	GAANACTTTTAATAAAGAAATCAGAAGATCCTTATGACGAAGATACGATACGTCCTTCTATC	296645
QY	541	ATTGACGACCAAAACGTAGSTAAATCAAGTTTATAGTAATGCTATTTTAGTGAAGATCGC	600
Db	296644	ATCGGTAGACCTTATGTTGGTAAATCTAGCTTGGTCANTGCTATTTTAGCGCAAGAACGT	296585
QY	601	GTTATCGTCTTAATGTTTGACGGGACAACGAGAGACGCTATTGATACAGAGTATAGTTAT	660
Db	296584	GTTATGTTGCTTAATGTTGCTGGTACAACCTCGAGATGCCATTTGATACCGAGTACTCTAT	296525
QY	661	GATGACAAGATTTATGTTTAAATCGATCTGCTGTATGCTTAAAAAAGGAAAAAGTATAT	720
Db	296524	GATGACAAGATTTATGTTATTCATTGATACCTGCTGGAATGAGAAAAAAGSTAAAGTGTAT	296465
QY	721	GAATCAACTGAGAAATATTCAGTATTAAGAGCTTTAAAGCGATTGAACGTTTCAATGTT	780
Db	296464	GAATCGACTGAAANAATTTCTGTATTACGTGCATTTAAACGGATTAGCGGTTCACAGANGTA	296405
QY	781	GTTTATGTTGTTATTGATGCAAGACAGAGGCATTATTGAACAAGATTAACCGTGTTCAGGA	840
Db	296404	GTTATGATAGTTTATCGATGCTGGAACAAGGTATTAATGACAAAGATTAACCGTGTAGCTGCG	296345
QY	841	TATGCACATGAACAAGGTAAAGCAGTCTGATTGTCGTAAATTAATGGGATACTCTGTGAA	900
Db	296344	TATGCACATGAGGAAGGTAAAGCTATTGTCATTGTAGTAAATAAATGGGATACAGTTGAA	296285
QY	901	AAAGATAGTAAACCGATCAAGAAATTTGAGAGTCACTAGCTAAAGAAATTCGAATTTTAT	960
Db	296284	AAAGATAGTAGACANTGAAANAATTCCTGATGATGTTAGAAANTGANTTTCAATTTTAT	296225
QY	961	GATTATGCACAAATTTGCTTTTGTCTGCTAAAGACGCACAAAGATTACGTACATTTATTC	1020
Db	296224	GATTATGCTCAATCGCTTCGTATCAGCAAAAAGAGGGGTAAAGATTTAAAAACATTTATTC	296165
QY	1021	CCTTACATTAATGAAGCAAGTGAAGAACCATAAAAACGTTGTCMAAGTTTCAACTTTAAAT	1080
Db	296164	CCTTATATCAATCAAGCCAGTGAANAATCATAAAAAGCGTGTCCAAAGTTCTACACTAAAT	296105
QY	1081	GAAGTCGTTACTGATGCCAATTTTCCATGACCCCTACACCACAGACAAAGGTAGACGTTTG	1140
Db	296104	GAAGTGTGTTACTGATGGCCATCTCTATGAATCCAAACACCTACTACAAAAGGTAGAAGACTT	296045
QY	1141	AATGTCCTTTTATGCAACAAAGTTGCTGTATAGAACCACCGACATTTGCTGTTATTTGTTAAT	1200
Db	296044	AATGTATCTTATACAACCTACAGTTTGCAAATTAAGAACCCGACATTTTGTACTATTTGTCAAT	295985
QY	1201	GATGTAGAAATTAATGCATTTTCTTATTAACCGCTATTTAGAGAAATCAAAATCCGTGCCGT	1260
Db	295984	GATGTTGAATTAATGCATTTTCTTATAGGAGATTTTTAGAAANTCMAATTCGTAATGCT	295925
QY	1261	TTTGGTTTTGAGGTTACACCAATTTTCATATTATAGCTCGAAAGAGA	1305
Db	295924	TTTGGTTTTGAGGAACACCTTATTCATATTATTTCCAAAGAAAGA	295880

RESULT 8

AE017002

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

gene

CDS

304680 bp DNA linear BCT 16-MAY-2003

Bacillus cereus ATCC 14579 section 5 of 18 of the complete genome.

AE017002 AE016877

AE017002.1 GI:29894935

Bacillus cereus ATCC 14579

Bacillus cereus ATCC 14579

Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group

1 (bases 1 to 304680)

Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelson,B., Kapatal,V., Bhattacharyya,A., Reznik,G., Mikhailova,N., Lapidus,A., Chu,L., Mazur,M., Goltsman,E., Larsen,N., D'Souza,M., Walunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fonstein,M., Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.

Genome sequence of Bacillus cereus and comparative analysis with

Bacillus anthracis

Nature 423 (6935), 87-91 (2003)

22608415

12721630

2 (bases 1 to 304680)

Candelson,B., Gailloux,K., Ehrlich,D.S. and Sorokin,A.

The number of ribosomal RNA operons in Bacillus cereus

Unpublished

3 (bases 1 to 304680)

Ivanova,N., Sorokin,A., Anderson,I., Galleron,N., Candelson,B., Kapatal,V., Bhattacharyya,A., Reznik,G., Mikhailova,N., Lapidus,A., Chu,L., Mazur,M., Goltsman,E., Larsen,N., D'Souza,M., Walunas,T., Grechkin,Y., Pusch,G., Haselkorn,R., Fonstein,M., Ehrlich,D.S.D., Overbeek,R. and Kyrpides,N.

Direct Submission

Submitted (12-MAR-2003) INRA, Genetique Microbienne, Domaine de

Vilvert, Jouy en Josas 78352, France

Location/Qualifiers

1..304680

/organism="Bacillus cereus ATCC 14579"

/mol\_type="genomic DNA"

/strain="ATCC 14579"

/db\_xref="ATCC:14579"

/db\_xref="taxon:226900"

68..427

/locus\_tag="BC1239"

68..427

/locus\_tag="BC1239"

/codon\_start=1

/transl\_table=11

/product="hypothetical Membrane Spanning Protein"

/protein\_id="AAP08224.1"

/db\_xref="GI:29894936"

/translation="MLRRKLLVFLTPVLAHLISMTKIELMALFLGYVFISNLNRI

OEQSLICFLFSISLEFLSIVSILLNELFLRIHSFTLMKPNWIIQAICTYVFVFL

KEIVGKOTVFQDNKWE"

551..2041

/locus\_tag="BC1240"

551..2041

/locus\_tag="BC1240"

/codon\_start=1

/transl\_table=11

/product="L-lactate permease"

/protein\_id="AAP08225.1"

/db\_xref="GI:29894937"

/translation="MLALPILNHWIFCLFKQTSRLSSLSIVASVIGVLLSPFQPL

GISTVNTATIKGHLICFVGYVFLGFIPLFMKMGYIDQAVFLFEVTDRLQML

LMKGFCLIESVSGFGIGVFWAFLPLSKFKFQVAVLSFGLLASSNGAVATGTL

IGSOLINHPITLITNTALLSPNPAFLVILSHVGVGMQAVIEKRWEGVFFLFLSI

GYLSYVAIVEVLGALISVITTFGFIILKKGSONLITEHAATREVSIIKIL

SPYLFATVCLISLVPALSDVFLVILDLKSYLELYSPGMGLWTCFLTFIF

FRIPNSIKQSLOTQKIPFAITTFMFIATSELMGASGHHSLAKTAGDTGTFVF

EVAPFIIGAIGSFTGNSAGNMTKLQMOANVALPHQVITTLONTASSVATIACT

```
gene
CDS
SRTLGCAYLCNIPYRENEELLKKTLLMIFGAVLLVVVEVIVYMLRN"
2132..2818
/locus_tag="BC1241"
2132..2818
/locus_tag="BC1241"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db_xref="GI:29894938"
/translacion="MTLKNYGVLTGTIOSEIGKGTTPHYOVHLOGETGVDRIRAINV
KSOYSEVLPFASNINKEAIHILPLTFPTFTFIRKNGRVALDTIRGNLFDNSQMI
PUPPEKAGNDLNEIKERYIKRAIEKARYIFAFGRMGPEENTPSYHFHFXPGNOMI
DHDGNGVNRWQGDGWDQGGILIHFEKEEEMIGIFLAFOSQSNCTDEEGHARVPV
EHCDFKGNIK"
complement(2830..3663)
/locus_tag="BC1242"
complement(2830..3663)
/locus_tag="BC1242"
/codon_start=1
/transl_table=11
/product="Integral membrane protein"
/protein_id="AAP08227.1"
/db_xref="GI:29894939"
/translacion="MELVNSIFQILLTSVQLLSLIGIIVIGFLGLYLESLRMVWS
RAGKRGLLTAMGVPVHELGHAIMCVLFRHKIVATQFPPTOTSOALGYVQHVNQ
KSVYORIGNFTIGIPITISGITALILMRYFVPSVPLENTLEKTIASTSYNEMIQ
NMLLSTFVLLSLFTIGLNLNPSFLFTAICTISAHIALSKPDIGSDGVNMFIV
FLFNLIITGLFOYDSNQLIGOVKNYNYLIAFSSVALLFSCISTLSVFSFYKIRGRS
F"
3860..4630
/locus_tag="BC1243"
3860..4630
/locus_tag="BC1243"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP08228.1"
/db_xref="GI:29894940"
/translacion="MKPCSCHTNAAEAKFCGNGHSLTEEVVASSQEGEPQAKEA
RNETVEAKFASGYPOFFKHALOAPTAIMKSGAEIVRNGIVSLVILGFLGACIFYR
NMSAASATRFAPDISPTFEGSIVTFEFLILFLVCFIIFVSCNMKMSFSFLE
VCGINGTATITVILVLSFLESFLLIFLIPILLVNAITIGISITVAVKLKONGLO
LVYTLIANVLGILFVPLWSYISTIIQALTKGITGF"
4650..5885
/locus_tag="BC1244"
4650..5885
/locus_tag="BC1244"
/codon_start=1
/transl_table=11
/product="IC hypothetical 16680"
/protein_id="AAP08229.1"
/db_xref="GI:29894941"
/translacion="MNVTCKGAQFEDGVQFCNQCCKRGRPVVKNKMSGGFKVGITL
LALFVILGILGASYSITUAVDRMITLGERGKELAEIVTADDPVMTRESL
PLFTSYNSVYNELKGYLRQEKDGIERVDFSLTKDGKFFEFIDRYKLAKTY
YTLTLNKGYSVNLKNGEKIDTKDKKFEKQYFPLPGNOVFQSDYKNDYVRLSREK
VYLMQSONYTIOLTLQGOYITVQTNAGTATLYVNOKPVTLAGEEITWGPVATDGS
ATYILERGESGRETTOYETITLSSYNLPFKKSTKTVVYVNTPTATGHVNGFI
FPDSIRKLTSGADLYLSKQOLKIARNEIYARHGHIFQTKMQAYFAKQSWYRNPYF
KGLTDITENIELIKSR"
complement(5931..6362)
/locus_tag="BC1245"
complement(5931..6362)
/locus_tag="BC1245"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP08230.1"
/db_xref="GI:29894942"
/translacion="MCAGLLAOLPDEPQEPKPKPAVSLDQEFINATYISARANLSN
DTEFAQDLTKATQAFASASHACKSIFPVGSVKIINFSDDTITVTASPESDHALKIIVPP
HSEALTSGPLSHQASTSGSSSKVTFPLPHIFSRDGPAAK"
```

```
gene
CDS
6560..7894
/locus_tag="BC1246"
6560..7894
/locus_tag="BC1246"
/EC_number="1.6.99.3"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase"
/protein_id="AAP08231.1"
/db_xref="GI:29894943"
/translacion="MNYVIIGDGAAGMSATHQIVTNDENANVYVLEKGETSYSAQCGL
PYYISGAIATSTKLTAKHNVKTRDQYIGIDAKVRHEVTKVDTEKMYVAEHTKTKDVE
FYDRLKATGVPVPMPEWREGLQVHLKTIPTDAERILKLTLEKNVEDVTYIGGGA
IGLEMAETVELGKRVATIERNDHIGTYIDADMAEYIHKEADKHIIILFNENKAFK
GNRVEAIEIDKGTAKADLVLSVSGVKPNTDFLECTNIRKNHKGAIENVAYMOTNVQD
VYAGDCATHYHVIKEIHIDHPLGTTANKQRLAGLANLMDKRRAFKCTGLCTGTLIKFD
LTLARTGLNEKEAKGLHPIYKTVKVDSTNMAGYVPNAKPLYLXLLYRSDTKQLLGGV
IGEEGVDKRIDVTAMALFNKMSIHLEDDVDLSYAPPYNSVWDPIQQAARAE"
8055..8465
/locus_tag="BC1247"
8055..8465
/locus_tag="BC1247"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP08232.1"
/db_xref="GI:29894944"
/translacion="MAINMKTIEMHIAESNARHEEDRGHVVEEMKEVCVGLDNATLIY
TKNPFCEGKKVVEFFQDHVYVIGQEKDEYVEIEKLYDAITNSLNATNDKNTTLEK
FANGQSINDSLNDNTGTNMLFARQIKSIFKLI"
complement(8496..10037)
/locus_tag="BC1248"
complement(8496..10037)
/locus_tag="BC1248"
/EC_number="1.6.99.3"
/codon_start=1
/transl_table=11
/product="NADH oxidase (NOXASE)"
/protein_id="AAP08233.1"
/db_xref="GI:29894945"
/translacion="MQLNFTFHLHLSFIDETMPNHNVDHKOADPLPKLTRNVPTIPLS
LEIPLSNSTPTPLNEIGHVLTSGVTOQLNSGRNVRLLRSIPSGALYPNELY
IYLDIDYDPGYHYDAAHRLILLREGNFOSYLADAGNRCNIHOCFGAAYVSTHEW
KNPFYNSFYLOGLDGVLIGQLLECAKOFYNTNGVYFQFLDRALNHLGLSVEE
SVYVPLSTETPTNFHNDYHKNKTVTSHELLQOIPPLAHNFVSKKHVDEYPMITK
INEASIMESTAFOTLAHEEHYQVHEHAVOLPKVERLSYDFLQLCRRKRYSPADPFLT
KMDAAELATLQEASLSLPYSNDLDGKYINENARVSLYCGFYNNVKDKNGAYTNSKT
HSTOPLRYGLRYPLOSSNTMDNVNLFQVSLCHVVCCKDYVTYNALGYRGYRTHOMEA
GILVHKLVPAATAMCMGHPHLLSFDNTNSCDQLYGIQAGNETSLIOVPVGAYRARNHLK
GALHN"
complement(10053..11987)
/locus_tag="BC1249"
complement(10053..11987)
/locus_tag="BC1249"
/codon_start=1
/transl_table=11
/product="hypothetical Cytosolic Protein"
/protein_id="AAP08234.1"
Query Match 56.2% Score 733.8; DB 1; Length 304680;
Best Local Similarity 72.6%; Pred No. 5.3e-118;
Matches 948; Conservative 0; Mismatches 357; Indels 0; Gaps 0;
QY 1 ATGACTAAACCTATAGTACTTATCTAGGTAGGCTAATAGTAGTAAATCTACAAATTTT 60
DB 244018 ATCCGAAACCAAGTAAATAGCTAGTGTCCCGCAACGATAGGAAATCTACTATTTT 111
QY 61 AATAGAAATAGTTGGAGAACGTGTTTCGATGTTGTAAGATACGCCAGGTGTAAACACAGAT 120
DB 244078 AATAGAAATGTTGGAGAAAGAGCTTCAATCGTAGAAGATATTCACGCTGTACACGCCGAT 244137
QY 121 CGTATTATTCTTCAGGTGAGTGGTTTACACATGATTTCATATTTTCAATATTATTAGTACAGTGGT 180
```



```

KGELEKNLPGLTVTIKMPFAQKLESDSGQFIDISFTGCGPDPDAITFELDMFITDGS
QNMGTSNREIOLINKAKQSDVKGWNNOLLAAKILFEDAAIAPVFORGRSILER
ETIKDVIHKYGEUSFWKASVEK"
complement(1956. .2099)
/locus_tag="BA1198"
complement(1956. .2099)
/locus_tag="BA1198"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP25160.1"
/db_xref="GI:30255151"
/translation="MIGVITNSIGEKLFQGGKMKFVNKSLDVSOGMNEVGGKGIYKK
RY"
2731. .3126
/locus_tag="BA1200"
2731. .3126
/locus_tag="BA1200"
/notes="similar to SP:P37687, PID:409796, GB:U00096,
PID:1790459, SP:P32687, PID:409796, GB:U00096, and
PID:1790459; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAP25161.1"
/db_xref="GI:30255152"
/translation="MVTLYSSPCTSCRAKAKLWLENNHIPVTERNIFSDPLTIEEIKE
ILRMTEGSDTEIISTRKSVFOELNVLNLESLPLQDLYKMRIDYGCILLRPMIMDEKRLQ
VGYNDEIBRFPLPRTVRFQLEAQRVFN"
complement(3176. .3850)
/locus_tag="BA1201"
complement(3176. .3850)
/locus_tag="BA1201"
/notes="identified by match to PFAM protein family HMM
PF03772"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AAP25162.1"
/db_xref="GI:30255153"
/translation="MDLEFLTSVMIVGIDVVGCGNAIVIALASRNLPDSKRNKAIL
IGTLAIVRIITLAVILQDIPFLQIGGVLLTIAVNLNTDNSDLSIOGKMTLL
FQNVRTVPAQVLMGFDNVAIAGAHRPSLVIIIGLISIPILMGSKMLMWERF
PFLYCGAAILAYTAGKMYTHEDRLATEFHNNPSFTAGIPLYFTILCTGFTVQOVR
LRNVKH"
4112. .4210
/locus_tag="BA1202"
4112. .4210
/locus_tag="BA1202"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP25163.1"
/db_xref="GI:30255154"
/translation="NICHMSGESLHNNSKQEGRVVFWILKELMIIR"
4207. .4860
/genes="meca"
/locus_tag="BA1203"
4207. .4860
/genes="meca"
/locus_tag="BA1203"
/notes="similar to GB:X03350, GB:X03350, GB:X03350,
GB:D00137, GB:M21692, GB:M24317, GB:M24308, GB:M24309,
GB:M24310, GB:M24311, GB:M24312, GB:M24313, GB:M24314,
GB:M24315, GB:M24316, GB:M37067, GB:X15447, GB:X15448,
GB:X15454, GB:X15450, GB:X15451, GB:X15452, GB:X15453,
GB:X15454, GB:X15455, SP:P00325, SP:P00326, SP:P07327,
PID:1723739, PID:178098, PID:178100, PID:178111,
PID:219428, PID:28416, PID:598143, PID:825622, GB:X03350,
GB:X03350, GB:X03350, GB:D00137, GB:M21692, GB:M24317,
GB:M24308, GB:M24309, GB:M24310, GB:M24311, GB:M24312,

```

```

GB:M24313, GB:M24314, GB:M24315, GB:M24316, GB:M37067,
GB:X15447, GB:X15448, GB:X15449, GB:X15450, GB:X15451,
GB:X15452, GB:X15453, GB:X15454, GB:X15455, SP:P00325,
SP:P00326, SP:P07327, PID:1723739, PID:178098, PID:178100,
PID:178111, PID:219428, PID:28416, PID:598143, PID:825622,
GB:X03350, GB:X03350, GB:X03350, GB:D00137, GB:M21692,
GB:M24317, GB:M24308, GB:M24309, GB:M24310, GB:M24311,
GB:M24312, GB:M24314, GB:M24315, GB:M24316,
GB:M37067, GB:X15447, GB:X15448, GB:X15"
/transl_table=11
/product="negative regulator of competence Meca"
/protein_id="AAP25164.1"
/db_xref="GI:30255155"
/translation="MKFFITYIDIEDGFNREEIWDRESEELFWPMDEARDHDOF
FIDGPLWIOQAVDKGIEYLVTKAELSDGOKLELPQGVDMIIIDPLDEGIESLFOOE
LVEVEEQAGTFNEDGTGFLIKFNDPEDIKSLHRLIFEDIKDELYSFEDRYVYV
EFDEVLDHDEEIDRILSVILEYGEESTLTIRHVSYGKQIVKHALETIRNPFAPK"
4933. .6477
/genes="cls-2"
/locus_tag="BA1204"
4933. .6477
/genes="cls-2"
/locus_tag="BA1204"
/notes="similar to SP:O66043, SP:O66043, SP:O66043,
SP:O66043, SP:O66043, and SP:O66043; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cardiolipin synthetase"
/protein_id="AAP25165.1"
/db_xref="GI:30255156"
/translation="MKNTUKLIFVLLFALFVSLRMFIDVAFYSDVIGIKDVSILGI
ISILTVSAPFLGCVGFLENRHSRSLTLMVLVIGFVPGFAYLLFGQNRKRMPQ
KKALEDQAPLOYKGHEDYEERSLRNHKKHQLFLADRLGALNISTOTETRTLNGD
ETQAILDLGKRAKHIMMEYIVVRDLKGTIKIDILKIQKSGEVVRFYLDVAGSFK
LSKSYTEELNDAGVIEFPFVPRFPLNDKINYNRHKIVIDIGNEGFVGGNIGDHY
LGRDKYFGWRDTHLYLRGEAVOSQLFLQDMWFYMTGEAVLAPYLQAKVAEGEHNG
GVOLVAGPDKNWTETKHLFYAMIASARKSIWIATPYEPDDLSLAKVAALAGIDY
RLMPSKPKDRTVYASRSYFPELLDAGVKIYEKGLPKHKKVIVSDSLASIGTANM
DMSEHLNFENNAFLYDTSIRKLQVDFKDDLESSEIHVDRFKRRLHRRIVESTYR
LLEPLL"
6539. .7803
/locus_tag="BA1205"
6539. .7803
/locus_tag="BA1205"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAP25166.1"
/db_xref="GI:30255157"
/translation="NIPFIARKENGEKIHLLDHDELLHMKRKRSPFCIACGKEV
LMRLGKQRHMFHAKKVDSCLAFTAEATSEYHMYKRELLYRMFKRQNVHVEHTEHLP
QQRPDIFIEHAGRAKIAIEYOCANLSIEQLCKRTYSYWGAGIOVWIIGGNOLKRSAY
WMKFSIMAFSLQSYPOPLLIFCQSKQSFMKCTFTITSFIASFHIIYLPETTFE
EMLFSVPVQREVLKRENRKRDYFRQNALPILNHNKSLRLLYLQKRCNPMPFSEI
GVPLPSFAFOTNSFQWAFYMKGICELAVGDCISLKYVCGYIKKVTKRMLPYFSO
HINKVAVTEYMTFLCYAGVLYKVGMYTYRKIRDAVPLKGTGCEIHKHDEICLAHLSLF
EAKFNMRGCKGDIKTDRGII"
7854. .9680
/genes="pepf-1"
Query Match 55.7% Score 727.4; DB 1; Length 290117;
Best Local Similarity 72.3% Pred. No. 6.9e-117;
Matches 944; Conservative 0; Mismatches 361; Indels 0; Gaps 0;
QY 1 ATGACTAAACCTATAGTAGCTATTGTAGTAGTGCCTTAATGTAGGTAGTAAATCTACAATTTT 60
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280374 ATGCCGAACCAAGTAGTAAGCAATAGTAGGCCGCCGCAAGTAGGGAATCTACTATTTC 280433
QY 61 AATAGATAGTTGTGAGAACCTGTTTCGATTGTTGAAGATACCCAGGTGTGAACAGAGAT 120
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

280434	ATAGAAATTTGGTGGAGACAGAGTTTTCAAATCTGTAGAAGATATTCAGAGTGTAAAGCCGGAT	280493	
QY	121	CGTATTTATTCCTTCAGGTGAGTGGTTTAACACATGATTTCAATATATTTATGTACAGTGGT	180
DB	280494	CGTATTTATTCAGCGCTGGAGATGGTTTAATCTGAATTTAACATATTTATGTACAGCGTGA	280553
QY	181	ATTGAATTTGGTGATGCAACCATTTCCAAACACAAATTTAGAGCGCAGGCAGAAATTCGCCATA	240
DB	280554	ATTGATATTTGGAGACAGCCATTTTTCACAAATTCGTCAACAGCTGAAGTAGCAAT	280613
QY	241	GATGAAGCGGATGTTATTTATTTATGGTTTAACGTGGGTGAAGGATTTGACACAAACCGAT	300
DB	280614	GATGAAGCAGATGTTATCATTTTATGACAATTCGTGACCGGTGAACAGCAGCCGAT	280673
QY	301	GAATGGTGGCTCAAAATTTTATACAAATCTAAAAAACCCGGTCGTATTAGCGGTTAACAA	360
DB	280674	GAAGAAGTTTGCAAAAATATATATCGTTTCAAAATAAACACAGTTGTACTTGCAGTAAATAA	280733
QY	361	GTAGATAATATGGAATTCGGTCACAGACGTGTATGATTTCTATTCTATTAGGATTTGGTGAA	420
DB	280734	GTGGACAATCCGGAATCCGCAGTGACATTTATGATTTCTATGCATTAGGATTTGGCGAG	280793
QY	421	CCGTATCCCAATATCAGGTCACATCGTTTAGGTCCTTGGTGACATTTGTTAGATGCAGTTGTT	480
DB	280794	CCATCCCAATCTCAGGTACACACGGATTAGGATTAGGTGACATTTATAGATGAAGCTGCA	280853
QY	481	TCCTATTTTGGTGAAGAGGAAGAAGATCCCTATGATGAAGATACAAATTCGACTATCCATT	540
DB	280854	CAGCATTTTCCAAAGATTTGAAGAAGACGGATATGATGAAGATACAAATTCGTTCTTTTA	280913
QY	541	ATTGGACGACCAACAGTAGGTTAAATCAAGTTTACTAAATGCTATTTTAGTGAAGATCGC	600
DB	280914	ATTGGCGTCCAAATGTAGGGAATCATCACTTCTTAANTGCTCTTCTTGGTCAAGAGCT	280973
QY	601	GTATCGTTTCTAATGTTTCGAGGACACAGACAGACCTATTGATACAGATATAGTTAT	660
DB	280974	GTAAATTTGTTAGTATGTAGCAGGAACAACCTCGTGATGCTGTCATACACCATATAGTAA	281033
QY	661	GATGCAACAAGATTATGTTTTTAATCGATACTGCTGGTATGCGTAAAAAGGAAAAAGTATAT	720
DB	281034	GATGGAAAAGATTATGTAATTTATCGATACAGCTGGTATGGTAAAAAAGGAAAAGTAT	281093
QY	721	GAATCAACTGAGAAATATTCAGTATTAAGAGCTTTTAAAGCGATTTCAACGTTTCAAAATGTT	780
DB	281094	GAAGTACAGAAAAGCTATAGTGTACTTCGTGCACCTAAGACGGATTTGAAGCTTCTGACGTT	281153
QY	781	GTTTTAGTGGTATTGATGCAAGACAGAGCATTTTCACACAGATAAAGCTGTTGCAGGA	840
DB	281154	GTTTTAGTCGTTTTAGACGGTGAAGAGGAATATTGACACAGATTAATAAATTCGCTGGA	281213
QY	841	TATGCACATGAACAAGGTAAAGCAGTCGTGATGTCGTAAATAAATGGGATACTGTGGAA	900
DB	281214	TACGCTCATGATTCAGGACAGAGCTGTGCTGATCGTTTGAACAAATGGGACGCAAGTAA	281273
QY	901	AAAGATAGTAAACCGATGAAGCAAAATTTGAAGATGAAGTACGTTAAAGATTTCCAAATTTTA	960
DB	281274	AAAGATGAGAAAACAATGAAAGCATTTGGAAGAAACATTCGTCTCATTTCCAAATTTTA	281333
QY	961	GATTATGCAAAATTTGCTTTTGTCTGTGCTTAAGAACGCACAGATTTACGTACATTTATTC	1020
DB	281334	GAGTAGCCACCGATTTGTTATCTTATGTCAAAAACGAGAAACGTCACACAAACGTTTAATA	281393
QY	1021	CCTTACATTTAAGACGAAGTGAAGAACCATTAATAAACGTTGTTCCAAAGTTTCAACTTAAAT	1080
DB	281394	CCGGTATCATGATGAGGTAAATGAAGAGCCATAGTATCCGTATCCAAACGAATGATTTAAAT	281453
QY	1081	GAAGTCGTTTACTGATGCAATTTTCCATGAACCCCTACACCAACACACAAAGGTAGACGTTTG	1140
DB	281454	GATGTAATATGAGATCGCGTACCGTGNATCCAACGCCAACGCAATACCGTAGTCGCTTG	281513
QY	1141	AATGCTCTTTATGCAACAAGTGGCTATAGAACCACCGACATTTGTTGTAATTTGTTAAT	1200
DB	281514	AAGATTTTCTATGCAACAAGTTCGCGTAAACACCAACACATTTGTTGTTGTTGTTAATC	281573

Oy	1201	GATGTAGAAATAATGCATTTTCTTATTATAACCGCTATTTAGAGAATCAAAATCCGTCGGCT	1260
Db	281574	TATCCGAAGACTATTCGACCTTCATCATAGCGCCTCTTAAAAAATCGTTTACGTGAATCA	281633
Oy	1261	TTTGTTTTGAAGGTACACCAATTCATATTAGTCGCAAGACA	1305
Db	281634	TTTCGGCTTTGTAGGAACGCCGATTACCATTTATCGTAGAGCAACA	281678
RESULT 10			
AL591981/c			
LOCUS	AL591981	347050 bp	linear BCT 06-JUN-2002
DEFINITION	Listeria monocytogenes strain EGD, complete genome, segment 9/12.		
ACCESSION	AL591981	AL591824	
VERSION	AL591981.1	GI:16411141	
KEYWORDS			
SOURCE	Listeria monocytogenes		
ORGANISM	Listeria monocytogenes		
REFERENCE	Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.		
AUTHORS	Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T., Charbit-Aubert, F., Couve, E., de Daruvar, A., Dehoux, P., Dommann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dessusmet, O., Entian, K.D., Fshih, H., Portillo, F.G., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, I.M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F., Kurapat, G., Madieno, E., Maitournam, A., Vicente, J.M., Ng, E., Nedjari, R., Nordström, S., Novella, S., de Pablo, B., Perez-Diaz, J.C., Percival, H., Remmel, B., Rose, M., Schlueter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J.A., Voss, H., Wehlend, J. and Coseart, P.		
JOURNAL	Comparative genomics of Listeria species		
MEDLINE	Science 294 (5543), 849-852 (2001)		
PUBMED	21537279		
REFERENCE	11679669		
AUTHORS	2 (bases 1 to 347050)		
TITLE	Glaser, P., Frangeul, L. and Rusniok, C.		
JOURNAL	Direct Submission		
	Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE		
COMMENT	E-mail: pglaser@pasteur.fr Phone: +33 1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.		

```

i. .347050
/organism="Listeria monocytogenes"
/mol_type="genomic DNA"
/strain="EGD-e"
/db_xref="taxon:1639"
complement(95..841)
/gene="lmo1688"
/gene="lmo1688"
CDS
/note="similar to glucosyl dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="CAC99766.1"
/db_xref="GI:16411142"
/db_xref="SPTREMBL:Q8Y6J6"
/translation="MNKVALVTGSSRGLEIAIAKAGFYDIANFSNRKKAEEVQ
QTEQLGRKCVIFKAVQGVKEVLEKFAVDGEFRGLDIFINNAASGVLRPLMELEES
HMDTMYNNAKALIFAGQEAALMORHQSCKIISLSIGSIYVLENTYTVGVSKAAVE
SLTRYLVALEAPFGIYAVSVGGILIEDALNHPNREELLKQAVSKTPAGRIEIPNDL
VNAVILFLASERADMTIRGQTLVDGEGTLLV"
complement(join(845..1942,1948..1953))
/gene="lmo1689"
CDS
/complete(845..1942)
/gene="lmo1689"
/note="similar to A/G-specific adenine glycosylase"
/codon_start=1
/transl_table=11
/protein_id="CAC99767.1"

```



```

/db_xref="GI:16411143"
/translaton="MKRKLRLTDETKTKTAQEQALVSVYEANKRVLPWRNRTGTPYRIM
VSEIMLQKQKVDYPIVPIFRPMTQPTMESFVNADEADILKAMEGLGYISVRANLQTA
MKQVADFGSEVPTDLTLLSLKGGVPTAGAILSIAYNOAEPADVGNVARIYARVLE
IGDIDINAKSTRIFEVLYQLIDKKNPAFNOGLMEIGALVCTPTKPKMLCLPQFPC
EAKHNGEVNYPVKIKKVMKTKELLSIIVISEDGKTAIEKRKPNGLLNNHMOFFPIE
ISKKNDEKVAQLQFLHNYGLEVLLEDEPIAHIKHVSHLVNKMDIRVAKLQSAIPNEN
WYFATPEEMKRLAFPVPYOKMQANKDFKE"
/gene
RBS
RBS
RBS
gene
RBS
CDS
/gene="lmo1688"
/complement(850..855)
/gene="lmo1688"
/complement(850..855)
/gene="lmo1689"
/complement(1948..1953)
/gene="lmo1689"
/join(2129..3123,3130..3147)
/gene="lmo1690"
/2129..2134
/gene="lmo1690"
/2143..3123
/gene="lmo1690"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC99768.1"
/db_xref="GI:16411144"
/db_xref="SPTREMBL:O8Y6J4"
/translaton="MDTGHVYMGIALGALATVDPVAGSSOAAIGTWTATIIIGSOIP
DIDVLKKNADYIRNRGITHSLPMLAIMPLLISSILYLFPAAATFIHLGLWPTFA
VGLHIFVDIYNAGYQAVRPFKETWVAFGFIINTDFMIFGSHVVAIAANMLGSPVLT
FVLYLILALYVAVRVTORMIKHAYONLIPDSEIIIASIHFFQNRVAVTTKDIYY
VGAFRANSIIEKEDRFPVDPNEIIRSAKDKNLAAFISFSYNNRIRIEKLDGTV
TFDLRTSRNGHFPFVAVVKLDDDKIVSSISGTMFSTFKLYKKLAPVSI"
/complement(join(3130..3147,3156..3617,3625..3630))
/gene="lmo1691"
/3130..3147
/gene="lmo1690"
/complement(3130..3147)
/gene="lmo1691"
/complement(3156..3617)
/gene="lmo1691"
/notes="similar to deoxyuridine triphosphate
nucleotidohydrolases"
/codon_start=1
/transl_table=11
/protein_id="CAC99769.1"
/db_xref="GI:16411145"
/db_xref="SPTREMBL:O8Y6J3"
/translaton="MKVRGFEVYVNEASRKFPEOTISLPIRGDKSAGYDFESNETVTI
VPEKHIFMTDINKSYMOEDEVLIYVRSSIGIKGLLNGTGLIDSSYSNPGNDGN
IGTAIRNSNEPYSIENGERRVAGQVQKYLVDITDIDIVANESRVGCVGSGTER"
/complement(3625..3630)
/gene="lmo1691"
/complement(3629..3862)
/gene="lmo1692"
/complement(3629..3862)
/gene="lmo1692"
/codon_start=1
/transl_table=11
/protein_id="CAC99770.1"
/db_xref="GI:16411146"
/db_xref="SPTREMBL:O8Y6J2"
/translaton="WITYLDELTAELVKNPHLDKEQALNMIEMLWSDFESSYAKAGY
PYRGPEYADYVRQOIERHGSFSLHQVRKDPNK"
/complement(join(3864..4673,4680..4685))
/gene="lmo1693"
/complement(3864..4673)
/gene="lmo1693"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC99771.1"
/db_xref="GI:16411147"

```

```

/db_xref="SPTREMBL:O8Y6J1"
/translaton="MKITSVCOOKKERYNIFIDEKYNFSDVEVLARYOLMGKVKVL
MDYINDEFEALYTKTOIKTAAINFLSHRRVSEKELRDLRKOEHPFAIDELIKKLAN
TKOAMKIMKRNKSAKMLQOKIITDLOKGFTESELAKAANAATEATSELDVAEAEIILQ
KQIKETMRKKNRYKPSIAKOKITITSLMKGFSYDTIQSYLTENEISPEEE"
/complement(3872..3877)
/gene="lmo1692"
/complement(3872..3877)
/gene="lmo1692"
/complement(4680..4685)
/gene="lmo1693"
/join(4757..5673,5681..5708)
/gene="lmo1694"
/4757..4762
/gene="lmo1694"
/4771..5673
/gene="lmo1694"
/notes="similar to CDP-bebequose synthase"
/codon_start=1
/transl_table=11
/protein_id="CAC99772.1"
/db_xref="GI:16411148"
/db_xref="SPTREMBL:O8Y6J0"
/translaton="MHILLTGATGFIGDHLVHELEKSDHELYILTROKLKNRNVHYI
EMLNEDLPNLEDPVDCITNLAGAGLMDERTYDRKKVIVNSRIEATSAALSIVKRM
KSKPKLWNASATGAVTSKSTIYLOTEENTYADNPLGKTVYENKTKTSAASDLGIRV
VYARFGLVLTNGSGSPVPEKLPOTYTGFRGNGRQMTSWIHVDVVAAILFIFDHEQ
INKVNTAPHPVQOEKKFAERLCKMKHKPKYKTPVKKIIFILGERANTILDSORAYP
EGLASNHFREFTELOEALDLDL"
/5681..5708
/gene="lmo1694"
/complement(join(5682..8291,8299..8304))
/gene="lmo1695"
/complement(5682..5709)
/gene="lmo1695"
/complement(5694..8291)
/gene="lmo1695"
/notes="similar to putative membrane proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC99773.1"
/db_xref="GI:16411149"
/db_xref="SPTREMBL:O8Y6I9"
/translaton="MKEKMQAYAVQKNSTWVKIVFITFVNAFVIFEIINIATGIDY
PSLKNLSQSPQIFITMFTVGLIAVTPMLLYDYIVIKLLPKGFSFSPSHVIAAGWINT
FTNIGFGVGLASLRASFYKGNASHKEILLALSKIALFLVSLGSIYGLVSLATLLIP
GFADHFVYNYMILLAGCLYPIILTYITKKSKSLFVDLPKIKRELTLLIISLLENGFAP
GCFALICTLAGEPVDIFKVPPLFVIASVIGIASMVGPGVGTDDVVMILGLSOLQVSOE
LALAHMFLYRIEYVYIIPVVGILLFFVQKAGKNQDPLEGLPLLEKLVKVAHRELVFY
GSGLLILSSNAPNAYIYHVPFLKIMPPNFLTQITITVAFGLLLGLARGIEKTKK
AYITIVVLGCAIFNTLARVFSMKQALFGLIVLGLFLARNEFIREKLYITWTSKVIID
SIIFIVLAGTIVIGIYNSPNIKHSKEIPDYLRISAEHLMLVGVGVGVIAVSVLIY
IYLSTTKKELGSPFEAVKVKHAKNGNEVSHFTMRDKLFLFAAEGLVFSYRIIA
DKMYINGEPTGNMDKMEAAIEEYMMNADRGYRPVYFVRGTMTIPLYLHDGDFDIKLG
EEGFVDQNFMTSGKKKGERALNNKLEREGYTFEIEPPENHDTMTTLARVSDSEMIL
GREEKGFSGLGFTOTYILEQAPIAIKANGEGTIVGFASMPSTYDMDTIDLMRYSKEA
PSGIMDFLNTLFEKAKEDGCFOTENAGMAPLANVGESYAFGLERLACLVYRSQGFY
GFKGLRNFKSKYVTEWQKFAVFRKSSIAFTMLQLMLVGLKGRPLANSQVVLDFPLE
EETKKPDSE"
/complement(8299..8304)
/gene="lmo1695"
/complement(8318..9007)
/gene="lmo1696"
/complement(8318..8992)
/gene="lmo1696"

```

```

Query Match 52.4%; Score 683.2; DB 1; Length 347050;
Best Local Similarity 70.2%; Pred. No. 2.9e-109;
Matches 916; Conservative 0; Mismatches 388; Indels 0; Gaps 0;
QY 1 ATGACTAAACCTATAGTAGCTATTGTAGGTAGGCCTAATGTAGGTAAATCTACAATTTT 60
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```







[illegible]











FEATURES  
source  
Location/Qualifiers  
1. 2662  
/organism="Lactococcus lactis subsp. cremoris"  
/mol\_type="genomic DNA"  
/strain="MC1363"  
/db\_xref="taxon:1359"  
/clone\_lib="MC1363 Random Sequence Tag Library"  
/note="Vector: pSGM2; Site\_1: SmaI; Library of  
chromosomal fragments of L.lactis strain MC1363 was  
prepared by partial AluI digestion or by sonication."  
BASE COUNT 871 a 506 c 513 g 772 t  
ORIGIN

Query Match 41.2%; Score 159.4; DB 28; Length 2662;  
Best Local Similarity 63.4%; Pred. No. 4.9e-32;  
Matches 244; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

OY 3 TCTTCTCCTTCCACCAAAATGAGAAACACTGCATCTAACAAAGTCACCAAGACCTAAA 62  
DB 2038 TCTTCCGATTTCTATTAGGAAGATTTTGAAACAATAGCATCAATGCCCTGTGCCAATC 1979

OY 63 CCATCTGACCTCGATATCGGATCGGTTCCACCAAAATCCCTAAATGAATAGAAATCATACACG 122  
DB 1978 CGCTCAACAGCTGATACAGATATGGATCCCGAGTCCCAAGTGAATAGAAATCAAAAT 1919

OY 123 TCTGTACCATTTCCATATATATCTACCTTTGTTAAACCGCTTAATACGACCGGTTTTTAGAT 182  
DB 1918 TCCATTCCAGCTTCCGGGTTGTCGACCTTTATGACGATAGAAATGACAGGTTTATCTGTA 1859

OY 183 TTGTATAAAATTTGAGCGACCATTTTCATCGCTTTGTGTCAATCCCTTCACGACGTTAAC 242  
DB 1858 CGATAAAGATATTAGCAACTGCTTCATCGCATCACTAATCTCTTTCATCAACACG 1799

OY 243 ATAAAAATTAATACATCGCTTCATCTATCGCATTTCTGCGCTCTAAATTTGTGTT 302  
DB 1798 ACAGCGATTAATACATCTGTTCCGTCATGCGCATTTCCGCTGCGCGCAATTCAGTC 1739

OY 303 TGGATGTGTCATCACCATAATTTCAATACCACTGTATCAATATATTTGAATCATGTGTT 362  
DB 1738 ATGAAAGTTTCATCGAAAGTTCAATCCCGGTCGATGATGTTGAATTTACGGGTT 1679

OY 363 AACCATTCACCTGAAGAATAAATAC 387  
DB 1678 AACCATTCCTGTCGCATACATAC 1654

RESULT 2  
AI390536/c 552 bp mRNA linear EST 15-MAR-2000  
LOCUS mu83b02.y1 Stratagene mouse melanoma (#937312) Mus musculus cDNA  
DEFINITION clone IMAGE:652107 5' similar to SM:YPHC-BACSU P50743 HYPOTHETICAL  
48.8 KD GTP-BINDING PROTEIN IN CMK-GPSA INTERGENIC REGION. ;, mRNA  
sequence.  
ACCESSION AI390536  
VERSION AI390536.1 GI:4216543  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 552)

TITLE The WashU-NCI Mouse EST Project 1999  
JOURNAL Unpublished  
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu  
This clone is available royalty-free through LML; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
This read is a RESEQUENCE of a previously sequenced mouse clone  
This read has been verified (found to hit its original self in the  
correct orientation)  
Possible reversed clone: similarity on wrong strand  
MGI:397955  
Seq primer: -40RP from Gibco  
High quality sequence stop: 420  
POLYA-No.

FEATURES  
source  
Location/Qualifiers  
1. 552  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone\_image:652107  
/tissue\_type="melanoma"  
/dev\_stage="M2 cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/clone\_lib="Stratagene mouse melanoma (#937312)"  
; Site\_1: EcoRI  
; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
dT. From M2 cells, a highly metastatic derivative of the  
K-1735 (mouse) melanoma. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

BASE COUNT 220 a 63 c 94 g 174 t 1 others

Query Match 24.1%; Score 93.4; DB 9; Length 552;  
Best Local Similarity 56.0%; Pred. No. 2e-14;  
Matches 200; Conservative 0; Mismatches 151; Indels 6; Gaps 1;

OY 17 ACCAAATGAGAAACAACTGCATCTAACAGTCACCAAGACCTAAACACCTGACCTGA 76  
DB 351 ATCTAACTCTTAATGTTTTCATCTATAATAATCAACACTCTCTTCATGAATAGCTGA 292

OY 77 TATCGGATACGGTTCACCAAAATCTAATGAATAGAAATCATACACCTCTGTACGCAATTC 136  
DB 291 AATTGGGAAAAATATGTTCCAAAACCTAGTGAATAAATTTGAAGATATC-----AAAAAATTT 238

OY 137 CATATTATCTACTTTGTTAAACCGCTAATACGACCGGTTTTTTTAGATTGTTATATAAATTTG 196  
DB 237 GTTTCCTCTAGCTTTATAGCTGCGATTAAAAACTTTTTTGTGTTGTTTTCGAAGCAATC 178

OY 197 ACGGACCATTTTCATCGCTTTGTCCTTTCACGACCGCTTAACCATATAAATAATAC 256  
DB 177 TATACAAATGATCTCTGATCTATTTCACTCTACCATCTACTATAAATAATATA 118

OY 257 ATCCGCTTCATCTATGGCGATTTCTGCGCTCGGCTCTAATTTGTTTGAATGTTGCAATC 316  
DB 117 ATTTGCTCTCTCTATTTGCTATTGAGTTTGAATTTCTAATTTGTTTCTTGAATGATTTGTT 58

OY 317 ACCAATTTCAATACCACTCTATCAATAATATTAATAATGTTTGAATCAATTCACCTACCC 373  
DB 57 TTCTATTTCATTCCTGCTGCTATCAGATAAATAATTCCTTCTCTGCTCATCAAC 1

RESULT 3  
AA215138/c 431 bp mRNA linear EST 03-FEB-1997  
LOCUS mu83b02.r1 stratagene mouse melanoma (#937312) Mus musculus cDNA  
DEFINITION clone IMAGE:652107 5' similar to TR:G1046032 G1046032 HYPOTHETICAL  
50.2 KD PROTEIN. ;, mRNA sequence.  
ACCESSION AA215138  
VERSION AA215138.1 GI:1814868  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 431)





QY 374 TGAAGAATAAA 384  
 Db 98 TTCTCCATAAA 88

## RESULT 5

CNS0106X/c

LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC  
 DEFINITION BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION

VERSION

AL098595

KEYWORDS

AL098595.1 GI:5610206

SOURCE

GSS.

ORGANISM

Drosophila melanogaster (fruit fly)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the European Drosophila Genome Project (EDGP) -  
 http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC  
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
 d'Etude du Polymorphisme Humain) with funding provided by a MRC  
 project grant. The DNA was prepared from embryos by Alain Bucheton  
 and Genevieve Payan. It has been constructed in the vector  
 pBelobAC11.

## FEATURES

source

BASE COUNT 258 a 107 c 60 g 175 t 501 others  
 ORIGIN

## Query Match

Best Local Similarity 12.0% Score 46.6; DB 29; Length 1101;

Matches 71; Conservative 155; Mismatches 159; Indels 0; Gaps 0;

QY 2 ATCTTCTCTCTTCCACCAAAATGAGAAACAATCATCTAACAAGTCACCAAGACCTAA 61

Db 974 MTHTHH 915

QY 62 ACCATGTGACCTGATATCGGATCGGTTCACCAATCTAATGAATGAATATCATACAC 121

Db 914 TMTHTHH 855

QY 122 GTCTGTACGATTTCCATATATATCTATCTTTGTTTACCCCTAATACGACCGGTTT 181

Db 854 YMTMTTCTTTTMTTTHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 795

QY 182 TTGTATATAAATTTGAGCGACCATTTTCATCGCTTTGTGTCAATCTTCCACGCGTTAAC 241

Db 794 TMTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTT 735

QY 242 CATATAAATAAATACATCGCTTCATCTATCGGATTTCTCGCTCGCTCTAATTTGTGT 301

Db 734 HTTATATATATATTTTMTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHT 675

QY 302 TTGGAATGGTCATACCAATTTCAATACCCCTGTATCAATATATTTGAATCATGTGT 361

Db 674 WMTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTTHTT 615

QY 362 TAACCATTACCTGAAGATAAATA 386  
 Db 614 MTHHAACACACMTWAKAAWAA 590

## RESULT 6

B2554672/c

LOCUS

DEFINITION

pacsl-60\_4804.y2 pacsl-60 Pseudomonas aeruginosa genomic clone

ACCESSION

B2554672

VERSION

B2554672.1 GI:37162489

KEYWORDS

GSS.

SOURCE

Pseudomonas aeruginosa

ORGANISM

Pseudomonas aeruginosa

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 787)

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

1..787

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="1-60"

/db\_xref="taxon:287"

/clone="pacsl-60\_4804"

/clone\_lib="pacsl-60"

/note="clinical isolate 1-60 Whole genomic shotgun library."

BASE COUNT 180 a 229 c 251 g 123 t 4 others

ORIGIN

Query Match

Best Local Similarity 11.8% Score 45.6; DB 29; Length 787;

Matches 129; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 78 ATCGGATACGGTTCACCAATCTAATGAATAGAAATACACGTCTGTACGATTTCC 137

Db 515 ATGGCGACGCTTGGCCAGCCAGCGCTTGACCTCGCGCAAGCAGATCGGATCG 456

QY 138 ATATTATCTACTTTGTTAAACCGCTAATACGACGGTTTTTTTATAGATTTTGTATAAATTTGA 197

Db 455 ATGTTGTCGCTTGTGTCGATCAGGAAGTTGCCCTTGTCCTGCCAGGTGCTCG 396

QY 198 GCGACATTTTCATCGCTTTGTGTCATCTCTTACGACAGCTTAACCATATAAATAAATAA 257

Db 395 CCGATCATCTGCGCGCGCGCGCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336

QY 258 TCCGCTTCATCTATGCGGATTTCTGCTCGCTCTAAATTTGTTTGGATTTGGTGCATCA 317

Db 335 TCCGCTTCCTCGATGGCTTGCACAGATTTCTCGGCAATTTTCGCGCATGATCGCTTCCTCA 276

QY 318 CCAATTTCAATACCACTCTGATCAATAA 345

Db 275 TCGCGCGAGATGCTCGGATTCGATGA 248

## RESULT 7

AY079759/c

LOCUS

DEFINITION

AY079759 Scripps Pier 745 bp DNA linear GSS 06-NOV-2002

AY079759

AY079759

AY079759

AY079759

AY079759

AY079759

AY079759

```

uncultured marine virus genomic clone SI051p1D7L, genomic survey
sequence.
ACCESSION AY079759.1 GI:24744504
VERSION AY079759.1
KEYWORDS GSS.
SOURCE uncultured marine virus
ORGANISM Viruses; environmental samples.
REFERENCE 1 (bases 1 to 745)
AUTHORS Breitbart M., Salanon P., Andresen B., Mahaffy J.M., Segall A.M.,
Mead D., Azan F. and Rohwer F.
TITLE Genomic analysis of uncultured marine viral communities
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (22), 14250-14255 (2002)
MEDLINE 22949988
PUBMED 12384570
COMMENT Contact: Rohwer F
San Diego State University
5500 Campanile Dr., San Diego, CA 92102, USA
Tel: 6195941336
Fax: 619595676
Email: forest@sunstroke.sdsu.edu
Class: shotgun.
FEATURES
    source
    1..745
    /organism="uncultured marine virus"
    /mol_type="genomic DNA"
    /db_xref="taxon:186617"
    /clone_lib="SI051p1D7L"
    /clone="SI051p1D7L"
    /community="Scripps Pier (La Jolla, CA) uncultured virus
    community"
    /note="Marine viruses were isolated from 200 liters of
    surface seawater using a combination of differential
    filtration and density-dependent gradient centrifugation.
    Linker-amplified shotgun libraries were constructed by
    randomly shearing the total marine viral community DNA,
    end-repairing, ligating dsDNA linkers to the ends, and
    amplifying the fragments using Vent DNA polymerase. The
    resulting fragments were ligated into the pSMART vector
    and electroporated into MC12 cells (Lucigen; Middleton, WI
    USA)."
BASE COUNT 167 a 221 c 226 g 127 t 4 others
ORIGIN
Query Match 11.5%; Score 44.4; DB 28; Length 745;
Best Local Similarity 48.7%; Pred. No. 0.34;
Matches 114; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
QY 86 CGGTTACCAAACTCTAATGAATAGAAATCATACAGCTGTGACGCATTTCCTATATTATC 145
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 745 CGGTTCCGCAAGCCGAGANANTAGCTTCCAGGATGCTGCTTTGGCGCTTCCCTTC 686
QY 146 TACTTTGTAAACCGCTAATACGACCGGTTTTTTAGATTGTATATAAATTTGAGCGACCATC 205
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 685 TGCCTGTTGTCACAGGACGACGCGGCACTTCGTCGCGCAGCCAGCGGCAATTC 626
QY 205 TTATCGGTTTGTGTCATCTTCACGACGCTTAACCATATAAATAAATCAATCGGCTTC 265
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 625 NTCGTCATCGTGTGATGCCAGCAGCGCATCGATGACGACATCGCGCGCTCGGCATC 566
QY 266 ATCATGCGGATTTCTCGCTGCTCTAATTTCTGTTTGGAAATGGTCATACC 319
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 565 GTCAATGCTGACTTCGGTCTGCTGGCATGCGCTCCCGCAGGAGGCGCTCATC 512

RESULT 8
CNS0000D1/c 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR01J16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL065414
VERSION AL065414.1 GI:4938827

```

```

KEYWORDS Drosophila melanogaster (fruit fly)
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammos in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
    source
    1..1101
    /organism="Drosophila melanogaster"
    /mol_type="genomic DNA"
    /db_xref="taxon:7227"
    /clone="BACR01J16"
    /clone_lib="RPCI-98"
    /note="end : TET3"
BASE COUNT 280 a 104 c 123 g 211 t 383 others
ORIGIN
Query Match 10.7%; Score 41.6; DB 29; Length 1101;
Best Local Similarity 13.8%; Pred. No. 2.1;
Matches 34; Conservative 128; Mismatches 84; Indels 0; Gaps 0;
QY 138 ATATTATCTACTTTCTTAACCGCTAATACGACCGGTTTTTATAGATTGTATAAATTGCA 197
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1099 MTNHHMYTNTNHHMYHTTTHHMTNHHMTNMTTTHHTTHHTMTTHTNMTNMTTHTNMT 1040
QY 198 GCGACCATTTCTACGCTTTGTGCAATCTTCACGCACGTTAACCATATAAATAAATAACA 257
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1039 MCHHHHHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMH 980
QY 258 TCCGCTTCTATCTATCGGCTTCTCGCTGCGCTTAATTTGCTTTGGGAATGGTCATCA 317
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 979 TTTTATATNHHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMH 920
QY 318 CCAATTTCAATACCACTGTATCAATATATATGAATCATGTCTTAACCATTCACCTGAA 377
| | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 919 WCTATCTATNHHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMHMMH 860
QY 378 GAATAA 383
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 859 MHTTAH 854

RESULT 9
AZ525216 653 bp DNA linear GSS 07-MAY-2001
LOCUS 241PBE05 Pb MBN #21 Plasmidium berghel genomic 3', genomic survey
DEFINITION sequence.
ACCESSION AZ525216
VERSION AZ525216.1 GI:13965843
KEYWORDS GSS.
SOURCE Plasmidium berghel
ORGANISM Plasmidium berghel

```



[illegible]

[illegible]

This clone is available at CHGC in Shanghai.

```

FEATURES
  source
    1..709
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="PHTBAE05"
      /tissue_type="hypothalamus"
      /dev_stage="Fetal"
      /lab_host="BM25.8"
      /clone_lib="PHTB"
      /note="Vector: pTriplex2; Site_1: sf1A; Site_2: sf1B1"
BASE COUNT      185 a      149 c      145 g      227 t      3 others
ORIGIN
Query Match      10.1% Score 39.2; DB 9; Length 709;
Best Local Similarity 56.1% Pred. NO. 8.3;
Matches 74; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 218 TGTCAATCCTTCACGACGTTAACCATATAAAATAATACATCCGCTTCATCTATGCGGAT 277
Db 247 TGTCAATCTAGTACTATATTAGTAGTGAACCAATACATCCCTTTCTTAIGTCTTG 306
QY 278 TTCTGCCCTCGCTCAATTTGCTTTGGGAATGGTCATCACCACCAATTTCAATACCACCTGT 337
Db 307 TTCTCCTTTTCATGTAATTTTGTCTGTATGGAGCACCACCTGATAGAGTTGGCATTGGA 366
QY 338 ATCAATAATATT 349
Db 367 AACATTTTATT 378

```

Search completed: September 12, 2003, 20:43:34  
 Job time : 1049.84 secs